

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2002, 07:45:45 ; Search time 13.25 Seconds
(without alignments)
1253.522 Million cell updates/sec

Title: US-09-645-192-2

Perfect score: 2389
Sequence: 1 MKIFRCYFKHTLQCKRFLIF.....DWITLPSEKLFMDNRLTTS 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	863.5	36.1	428	1 G6NT_HUMAN	Q02742 homo sapien
2	848.5	35.5	428	1 G6NT_MOUSE	Q09322 mus musculu
3	840	35.2	427	1 G6NT_BOVIN	Q92180 bos taurus
4	730.5	30.6	400	1 BGIB_MOUSE	P97402 mus musculu
5	711	29.8	400	1 BGIB_HUMAN	Q06430 homo sapien
6	715	4.9	895	1 YAS5_HUMAN	Q58454 metanococ
7	108.5	4.5	661	1 WHI1_YEAST	P34761 saccharomyc
8	102.5	4.3	654	1 FGR2_HUMAN	Q01442 homo sapien
9	102.5	4.3	821	1 FGR2_HUMAN	P21802 homo sapien
10	102.5	4.3	821	1 FGR2_HUMAN	P21803 mus musculu
11	101.5	4.2	1442	1 DPO3_UREPA	Q996b4 ureaplasma
12	101.5	4.2	2054	1 YCF2_MOUSE	P09376 nicotiana t
13	101.5	4.2	2280	1 YCF2_TOBAC	P09376 nicotiana t
14	100.5	4.2	2607	1 BACB_BACLI	O68007 bacillus t
15	99.5	4.2	666	1 VOO1_VACCC	P21093 vaccinia v
16	99.5	4.2	823	1 CEK3_CHICK	P18461 gallus gall
17	99	4.1	398	1 A23D_DROME	Q24093 drosophila
18	97.5	4.1	934	1 SYL1_SUI50	P58176 sulfolobus
19	97.5	4.1	1314	1 TETX_CLOTE	P04558 clostridium
20	97	4.1	340	1 YM2_ARCFU	Q28662 archaeoglob
21	97	4.1	440	1 GAAP_HUMAN	Q00561 homo sapien
22	97	4.1	1036	1 YOB6_CAEEL	Q09559 caenorhabd
23	95.5	4.0	936	1 MSB4_HUMAN	Q15457 homo sapien
24	95	4.0	701	1 SYGB_HELPJ	Q924m9 helicobacte
25	94	3.9	364	1 DP3B_MYCCE	P47247 mycoplasma
26	94	3.9	454	1 DMA_BUCAP	P29434 buchnera ap
27	94	3.9	1344	1 IFN3_MOUSE	P23116 mus musculu
28	94	3.9	1675	1 CLH1_HUMAN	Q00610 homo sapien
29	94	3.9	1675	1 CLH1_BOVIN	P49951 bos taurus
30	94	3.9	1675	1 CLH1_RAT	P14142 rattus norv
31	94	3.9	5255	1 BACA_BACLI	O68006 b bacillrac
32	93.5	3.9	884	1 SECA_PORPU	P51381 porphyra pu
33	93.5	3.9	2787	1 TEL1_YEAST	P38110 saccharomyc

34	93	3.9	503	1 CPV1_SHEEP	O9X528 ovis aries
35	93	3.9	520	1 Y208_METTA	O60270 methanococ
36	92	3.9	908	1 DPO1_BORBU	O51496 borrelia bu
37	92	3.9	1103	1 CYGF_BOVIN	O02740 bos taurus
38	92	3.9	1382	1 IFB3_HUMAN	O14152 homo sapien
39	92	3.9	2022	1 ANM1_ONCVO	P21245 onchocerca
40	91.5	3.8	720	1 SPOT_MYCCE	P47520 mycoplasma
41	91.5	3.8	757	1 FRS1_YEAST	P38903 saccharomyc
42	91.5	3.8	1956	1 ARX1_PLAAT	O04956 plasmodium
43	91	3.8	344	1 Y613_METTA	O68030 methanococ
44	91	3.8	458	1 YC13_EUGGR	P05728 euglena gra
45	91	3.8	490	1 MET3_SCHPO	P78937 schizosacch

ALIGNMENTS

```

RESULT 1
ID G6NT_HUMAN STANDARD: PRT: 428 AA.
AC Q02742;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, last sequence update)
DT 01-NOV-1997 (Rel. 35, last annotation update)
DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-
DE ACETYLGLUCOSAMINYLTTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)
DE (CORE2-GLCNAC-TRANSFERASE) (C2GNT) (CORE 2 GNT).
GN GCNT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93028457; PubMed=1329093;
RA Bierhuizen M.F.A., Fukuda M.;
RT "Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta
RT 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6-GlcNAc transferase by gene
RT transfer into CHO cells expressing polyoma large tumor antigen.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9326-9330(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96078409; PubMed=759796;
RA Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
RT "Genomic organization of core 2 and I branching beta-1,6-N-
RT acetylglucosaminyltransferases. Implication for evolution of the
RT beta-1,6-N-acetylglucosaminyltransferase gene family.";
RL Glycobiology 5:417-425(1995).
CC - FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.
CC - CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
CC GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R = UDP + BETA-D-
CC GALACTOSYL-1,3-N-ACETYL-D-GLUCOSAMINYL-R.
CC - PATHWAY: GLYCOSYLATION.
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC - TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ACTIVATED T-LYMPHOCYTES
CC AND MYELOID CELLS.
CC - SIMILARITY: TO I-BRANCHING ENZYME (IGNT).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: M97347; AAA35919.1;
DR EMBL: L41415; AAA96661.1;
DR PIR: A46293; A46293.
DR MIM: 600391;
DR InterPro: IPR003406; Branch.

```


DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-
 DE ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)
 DE (CORE2-GLCNAC-TRANSFERASE) (C2GNT) (CORE 2 GNT).
 GN GNTL.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Lung;
 RC Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RL Li C.M., Cheng P.W.;
 CC -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
 CC GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYLTRANSFERASE (I-BRANCHING ENZYME)
 CC GALACTOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYLTRANSFERASE)
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
 CC -1- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.sib.ch).
 CC -----
 DR EMBL: U41320; AAA83244.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 KM Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KM Golgi stack; Glycoprotein.
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 10 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 33 427 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 427 AA; 49733 MW; 41B9CBFD948D4196 CRC64;

Query Match 35.28; Score 840; DB 1; Length 427;
 Best Local Similarity 42.88; Pred. No. 1,4e-54;
 Matches 185; Conservative 73; Mismatches 134; Indels 40; Gaps 12;

28 LKLTANVRL--FQKDIYL-VEXSLSTSPFVRNRYHVKDE-VRY-----EV 71
 1 LKRLMKRLKLFSPFTKYFLFLASVYTFYLR--IHQTEFNPCHLLFEPNPSNI 57
 72 NCSGIYQDEPLEJGK----SLEIRRDIIIDEDDDVVAWTSDDIYTLRGVQKYSKE 127
 58 NCTILLOGDVEIQKYLESLTYVFKFRARWTINDYINMGDCASFKKKRYITPEPSKE 117
 128 EKSPFIAYSLTVVHKDAIIVERLIIAHINOHNYICIHDRKAPDFEKYAMNNLKCFENIF 187
 118 EAGPPIAKSYIVHHKIKEMDLRLAIYMPONFYCIHDANKSEKSFLLAAVGIASCFSNVF 177
 188 IASHLEAVEVAHISRLQADLNCISDLKSSIQWKYVINLGGDFPLKSNFEVSELKLN 247
 178 VASQLESVYVASRVOADLNCMODLYOMNAGWKYLLINLCGMPDPITKTNLEIVAKLELM 237
 248 GANNLEIVKPPNSKLEPFTYHHELRVPYEVYKLPITNT--SKGAPPHNIQIFVGSAVF 305
 238 GENNLETEKMPSEHKERRKHH-----YEVVNGKL-TNMGTDIHPLETPPLSGSAHF 289
 306 VLSQAFVYKIFPNSIYQDFPAMSKDYSPDEHFWATLIRVPGIIGELISRSAG--DVSLOS 364
 290 VYSREYVEYVQNONIQNFEMEMAKDYSPEDEIWTATIORIPEVVGSLSYKVDTSMDQA 349

QY 365 KTRIVKWNYYEGF-----YPSCTGSHLRVSYGAGAEIRWLKNGHEANKFSKVPDI 419
 DB 350 IARFVKWQIFESDVGSKADYPPCS-VHVASVCFSGAGDLNHLVHLEANKFDTIDLF 408
 QY 420 LKICLAERKLEED 431
 DB 409 AIQCDEHLRHK 420

RESULT 4
 BGIB_MOUSE
 ID BGIB_MOUSE STANDARD: PRT; 400 AA.
 AC P97402;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE N-ACETYLGLUCOSAMINID BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE
 DE (EC 2.4.1.150) (N-ACETYLGLUCOSAMINYLTRANSFERASE) (I-BRANCHING ENZYME)
 DE (IGNT) (LARGE I ANTIGEN-FORMING BETA-1,6-N-
 DE ACETYLGLUCOSAMINYLTRANSFERASE).
 GN GNT2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97280061; PubMed=9134435;
 RA Magnet A.D., Fukuda M.;
 RT Expression of the large I antigen forming beta-1,6-N-
 RT acetylglucosaminyltransferase in various tissues of adult mice.";
 RL Glycobiology 7:285-295(1997).
 CC -1- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED
 CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I
 CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED
 CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL
 CC -1,4-N-ACETYL-D-GLUCOSAMINYLTRANSFERASE (I-BRANCHING ENZYME)
 CC -1,4-N-ACETYL-D-GLUCOSAMINYLTRANSFERASE (I-BRANCHING ENZYME)
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
 CC -1- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.sib.ch).
 CC -----
 DR EMBL: U68182; AAB39621.1; -
 DR MGD: MGI:1100870; Gcnt2.
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 KM Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KM Golgi stack; Glycoprotein.
 FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 7 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 26 400 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 400 AA; 45507 MW; 193F3F8823B9691 CRC64;

Query Match 30.66; Score 730.5; DB 1; Length 400;
 Best Local Similarity 40.08; Pred. No. 1.5e-46;
 Matches 156; Conservative 51; Mismatches 160; Indels 23; Gaps 6;

```

QY 43 IYLVEYSLSSTSEFVNRNRYTHVKEDEVYEVNCSGIYEDEPLEIGKSLERRDDIDLEDD 102
D 18 VCVYVYSLFSGGDQSYOKLNTISDSVRLSYOVSSPFIID-----GKSRFLMRKLL----- 64
QY 103 VYAMMSDDDIYOTLNGYAKOKLVSKREKSPFPAISLYVKADLIMVRLHAIYNOHNIYCI 162
D 65 MHKRPSCIEYVYOSHYYTAPLSOEQGFPLAYVYVHHNDPFRALFRALFMPQNIYCV 124
QY 163 HYDRKAPDTFYKAMNNLAKCFESNIFASKLEAVEYAHISRLQADLNCISDLKSSIQWKY 222
D 125 HYDEKATAEFKGAVEQVLSCFPNVLASKMEPVYGGISRLQADLNCISDLKSSIEVPKY 184
QY 223 VNLGGQDFPLKSNELYSCLKLNGAMLETVPKPSKLERFY-HHELRRVPEYVKL 281
D 185 AINTGQGFPLKTNKEIYQYLGKMKLTPGVLPFAHISRTRYVREHLSKELSYV-- 242
QY 282 PIRNISKEAPPNHIQIYVGSAYFVLSQAFKVIYFNNSIYODFPAWSKDTSPDEHFMAT 341
D 243 -IRITALKPPPHNLITIFGSAVYALSREFANFVLRDRAVDLHMSKDTSPDEHFMAT 301
QY 342 LIRVGEIGELSRSAQVSDLSQSTRLVKMYEGEFPYPSCTGSHLSVCITYGAELRML 401
D 302 LNRIGVGVSPNPNAS-----WTGNLRAYKMMDMEA-KHGCGCHYVHGICITYGMDLQWL 355
QY 402 IKDGHMFANKPKDVPILIKCLAEKLEEO 431
D 356 INSQSLFANKFELNTPYLPVECLELRLKER 385

RESULT 5
BGIR_HUMAN STANDARD; PRT; 400 AA.
AC 006430:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE N-ACETYLGLUCOSAMINIDE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE
DE (EC 2.4.1.150) (N-ACETYLGLUCOSAMINYLTRANSFERASE) (I-BRANCHING ENZYME)
DE (IGMT).
GN GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=9319405; PubMed=8449405;
RA Bierihuizen M.F.A., Mattei M.-C., Fukuda M.;
RT "Expression of the developmental I antigen by a cloned human cDNA
RT encoding a member of a beta-1,6-N-acetylglucosaminyltransferase gene
RT family.";
RL Genes Dev. 7:468-478(1993).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=96078409; PubMed=759796;
RA Bierihuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
RT "Genomic organization of core 2 and I branching beta-1,6-N-
RT acetylglucosaminyltransferases. Implication for evolution of the
RT beta-1,6-N-acetylglucosaminyltransferase gene family.";
RL Glycobiology 5:417-425(1995).
RN 13
RP TISSUE SPECIFICITY.
RX MEDLINE=98070745; PubMed=940506;
RA Sasaki K., Kurata-Miura K., Ujita M., Angata K., Nakagawa S.,
RA Sekine S., Nishi T., Fukuda M.;
RT "Expression cloning of cDNA encoding a human beta-1,3-N-
RT acetylglucosaminyltransferase that is essential for poly-N-
RT acetylglucosamine synthesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:14294-14299(1997).
CC -1- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED
CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I
CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED

```

```

CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL
CC -1,4-N-ACETYL-D-GLUCOSAMINYLR-UDP + N-ACTYL-BETA-D-
CC GLUCOSAMINYLR-1,6-BETA-D-GALACTOSYL-1,4-N-ACTYL-D-GLUCOSAMINYLR.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -1- TISSUE SPECIFICITY: IN THE ADULT, HIGHLY EXPRESSED IN PROSTATE AND
CC TO A LESSER EXTENT IN SMALL INTESTINE AND COLON. BARELY DETECTED
CC IN HEART, BRAIN, KIDNEY AND PANCREAS. NO EXPRESSION IN PLACENTA,
CC LUNG, LIVER, SKELETAL MUSCLE, SPLEEN, THYMUS, TESTIS, OVARY AND
CC PERIPHERAL BLOOD LEUKOCYTES. IN FETUS, HIGHLY EXPRESSED IN BRAIN
CC AND TO A LESSER EXTENT IN LUNG AND KIDNEY. BARELY DETECTED IN
CC LIVER.
CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION INCREASES DRAMATICALLY DURING
CC DEVELOPMENT AND ONCOGENESIS.
CC -1- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EXBL: L19659; AAB1777.1; -
CC DR EXBL: L19550; CAA79610.1; -
CC DR EXBL: L41607; AAA99832.1; -
CC DR EXBL: L41605; AAA99832.1; JOINED.
CC DR EXBL: L41606; AAA99832.1; JOINED.
CC PIR: A46297; A46297.
CC MIM: 600429; -
CC MIM: 110800; -
CC DR InterPro: IPR003406; Branch.
CC DR Pfam: PF02485; Branch.
CC DR Transfaser: Glycosyltransferase; Transmembrane; signal-anchor;
CC KMW Goldy stack; Glycoprotein.
CC FT DOMAIN 1 6
CC FT TRANSMEM 1 25
CC FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC FT (POTENTIAL).
CC FT CYTOPLASMIC (POTENTIAL).
CC FT LOMENAL, CATALYTIC (POTENTIAL).
CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 37 400
CC FT CARBOHYD 255 255
CC FT CARBOHYD 314 314
CC FT CARBOHYD 388 388
CC FT N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 400 AA; 45854 MW; 1469628690A1F43B CRC64;

Query Match 29.8%; Score 711; DB 1; Length 400;
Best Local Similarity 43.7%; Pred. No. 3.9e-45;
Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

QY 108 SPDCDIYOTLNGYAKOKLVSKREKSPFPAISLYVKADLIMVRLHAIYNOHNIYCHYDRK 167
D 70 SSKREYIYOSHYYTAPLSOEQGFPLAYVYVHHNDPFRALFRALFMPQNIYCVHVDK 129
QY 168 APDTFYKAMNNLAKCFESNIFASKLEAVEYAHISRLQADLNCISDLKSSIQWKYVINC 227
D 130 ATTEFKDAVYQVLSCFPNVLASKMEPVYGGISRLQADLNCISDLKSSIEVPMYVINC 189
QY 228 GQDFPLKSNELYSCLKLNGAMLETVPKPSKLERFY-HHELRRVPEYVKLPIRN 286
D 190 GQDFPLKTNKEIYQYLGKMKLTPGVLPFAHISRTRYVREHLSKELSYV--IRTT 246
QY 287 ISKEAPPNHIQIYVGSAYFVLSQAFKVIYFNNSIYODFPAWSKDTSPDEHFMATLIRV 346
D 247 ALKPPPHNLITIFGSAVYALSREFANFVLRDRAVDLHMSKDTSPDEHFMATLIRP 306
QY 347 GIGELISRSQVSDLSQSTRLVKMYEGEFPYPSCTGSHLSVCITYGAELRMLIKDGH 406
D 307 GVGSPNPNAS-----WTGNLRAYKMMDMED-RHGCGCHYVHGICITYGMDLQWVNSPS 360
QY 407 WFAKFDKVDPIILIKCLAEKLEEO 433

```



```

Db      361 LEFANKPELNTYPLVECL--ELRRER 385
|||||: |:|:|: |:|:|
RESULT 6
ID      VAS4_METJA      STANDARD:      PRT:      895 AA.
AC      058454:
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DE      20-AUG-2001 (Rel. 40, Last annotation update)
DE      HYPOTHEMETICAL PROTEIN MJ1054 (EC 1.1.1.-).
GN      MJ1054.
OS      Methanococcus jannaschii.
OC      Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OX      NCBI_TaxID=2190;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX      MBDLINE=96337999; PubMed=8688087;
RA      Bolt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA      Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.L.,
RA      Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA      Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA      Uitterlidge T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA      Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA      Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RT      "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT      jannaschii."
RL      Science 273:1058-1073(1996).
CC      -1- SIMILARITY: BELONGS TO THE UDP-GLUCOSE/GDP-MANNOSE DEHYDROGENASES
CC      FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U67548; AAB99056.1;
DR      TIGR: MJ1054;
DR      InterPro: IPR003586; Hmnc.
DR      InterPro: IPR003587; Hmnc.
DR      InterPro: IPR002025; NAD_Binding.
DR      InterPro: IPR001732; UDPG_MGDP_dh.
DR      Pfam: PF00984; UDPG_MGDP_dh; 1.
DR      SMART: SM00305; Hmnc; 1.
DR      SMART: SM00306; Hmnc; 1.
KW      Hypothetical protein; Oxidoreductase; NAD; Complete proteome.
FT      NE_BIND 2 19 NAD (POTENTIAL).
FT      ACT_SITE 261 261 BY SIMILARITY.
SQ      SEQUENCE 895 AA; 102614 MW; B511A68CCA41445 CRC64;
SQ
Query Match 4.9%; Score 116; DB 1; Length 895;
Best Local Similarity 20.1%; Pred. NO. 0.59; Indels 136; Gaps 23;
Matches 91; Conservative 73; Mismatches 153;

```

```

QY      198 AHISRLQADLNLDSL-----KSIQMKVYVLCGDDPPLKSNFELVSLKTKNGAN 250
DB      541 ENI--LNCGINCYNNKIPQMFNAKEIKWF-----LKGLFGDGIYVLRNNK 588
QY      251 MLETKPPNSKLEFRTYHHELRVPEYVKLPRTNISKEAPHNIOIFVGSAYFLVSOA 310
DB      589 -----NLNTEPAT-----VSKKM-AHSLILLQLGIYAS-- 617
QY      311 FVKYIFNNSIVODFRAMSKDTYSPDEHFWATLIRYGP-----GEISRSQDYSVD 361
DB      618 -VKKCYNN-----KST-----TWAYIRINGLEQYVKIGELFGKWKENYKDIAE 660
QY      362 LOSKT-----RLKMYVEGFYFPSCGSHLRVCIYGAELRWLIK 403
DB      661 SYKRNIPELGYKSKDNFALLEVKEIK-ETHSGYVYSVETENSL-LISYGLILNCPK 718
QY      404 DGHWFANKFD-SKVDPILIKLAELKEQQRDM 435
DB      719 DVKALIKQFENNINIEPILIKA-TDIYNEQIKW 750
RESULT 7
WH13_YEAST
ID      WH13_YEAST      STANDARD:      PRT:      661 AA.
AC      P34761;
DT      01-FEB-1994 (Rel. 28, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DE      01-OCT-1996 (Rel. 34, Last annotation update)
DE      WH13 PROTEIN
GN      WH13 OR YN197C OR N1382.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RT      Nash R.S., Furcher A.B.;
RL      Submitted (Aug-1993) to the EMBL/Genbank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=5286C / FY1679;
RX      MEDLINE=95242839; PubMed=7725799;
RA      Jondiaux J.-L., Coster F., Purnelle B., Goffeau A.;
RT      "A 21.7 kb DNA segment on the left arm of yeast chromosome XIV
RT      SSI and 8 new open reading frames of unknown function."
RL      Yeast 10:1639-1645(1994).
CC      -1- FUNCTION: INVOLVED IN SIZE CONTROL; POSSIBLE RNA BINDING PROTEIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U01095; AAA03320.1;
DR      EMBL: X78898; CA55511.1;
DR      EMBL: Z71473; CA96092.1;
DR      HSSP: P09012; WH13.
DR      SGD: S0005141; WH13.
DR      InterPro: IPR00504; WH13.
DR      Pfam: PF00076; rrm; 1.
DR      SMART: SM00360; RRM; 1.
DR      PROSITE: PS50102; RRM; 1.
KW      RNA-binding; 247
FT      DOMAIN 661 AA; 71253 MW; 47AB59716CF96043 CRC64;
SQ      SEQUENCE 661 AA; 71253 MW; 47AB59716CF96043 CRC64;
SQ
Query Match 4.5%; Score 108.5; DB 1; Length 661;

```


RN [1]
 RA SEQUENCE FROM N.A.
 RP TISSUE-Neonatal brain stem;
 RC MEDLINE=90360977; PubMed=1697263;
 RA Dione C.A., Crumley G.R., Bellot F., Kaplow J.M., Searfoss G.,
 RA Rita M., Burgess W.H., Jaye M., Schlessinger J.;
 RT "Cloning and expression of two distinct high-affinity receptors
 RT cross-reacting with acidic and basic fibroblast growth factors.";
 RL EMBO J. 9:2685-2692(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91045961; PubMed=2172978;
 RA Housaint E., Blauquet P.R., Champion-Arnaud P., Geanel M.C.,
 RA Torriglia A., Courtois Y., Breathnach R.;
 RT "Related fibroblast growth factor receptor genes exist in the human
 RT genome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8180-8184(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91274356; PubMed=1647213;
 RA Seno M., Sasada R., Watanabe T., Ishimaru K., Igarashi K.;
 RT "Two CDNA's encoding novel human FGF receptor.";
 RL Biochim. Biophys. Acta 1089:244-246(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Stomach cancer;
 RX MEDLINE=90332706; PubMed=2377625;
 RA Hattori Y., Odagiri H., Nakatani H., Miyagawa K., Naito K.,
 RA Sakamoto H., Katoh O., Yoshida T., Sugimura T., Terada M.;
 RT "K-sam, an amplified gene in stomach cancer, is a member of the
 RT heparin-binding growth factor receptor genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:5983-5987(1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92212948; PubMed=1313574;
 RA Katoh M., Hattori Y., Sasaki H., Tanaka M., Sugano K., Yazaki Y.,
 RA Sugimura T., Terada M.;
 RT "K-sam gene encodes secreted as well as transmembrane receptor
 RT tyrosine kinase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:2960-2964(1992).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE=93016048; PubMed=1400433;
 RA Dell K.R., Williams L.T.;
 RT "A novel form of fibroblast growth factor receptor 2. Alternative
 RT splicing of the third immunoglobulin-like domain confers ligand
 RT binding specificity.";
 RL J. Biol. Chem. 267:21225-21229(1992).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Blood;
 RA Steinberger D., Mueller U.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP VARIANTS CS HIS-340; ARG-342; SER-342; TYR-342 AND CYS-354.
 RX MEDLINE=95078932; PubMed=7987400;
 RA Reardon W., Winter R.M., Rutland P., Pulleyn L.J., Jones B.M.,
 RA Malcolm S.;
 RT "Mutations in the fibroblast growth factor receptor 2 gene cause
 RT Crouzon syndrome.";
 RL Nat. Genet. 8:98-103(1994).
 RN [9]
 RP VARIANTS CS CYS-328 AND CYS-347, AND VARIANT JWS GLY-344.
 RX MEDLINE=95179174; PubMed=7874170;
 RA Jabs E.W., Li X., Scott A.F., Meyers G., Chen W., Eccles M., Mao J.,
 RA Charnas L.R., Jackson C.E., Jaye M.;
 RT "Jackson-Weiss and Crouzon syndromes are allelic with mutations in
 RT fibroblast growth factor receptor 2.";
 RL Nat. Genet. 8:275-279(1994).
 RN [10]
 RP VARIANTS CS.
 RX MEDLINE=95384152; PubMed=7655462;

RA Oldridge M., Wilkie A.O.M., Stanley S.F., Poole M.D., Pulleyn L.J.,
 RA Rutland P., Hockley A.D., Wake M.J.C., Goldin J.H., Winter R.M.,
 RA Reardon W., Malcolm S.;
 RT "Mutations in the third immunoglobulin domain of the fibroblast growth
 RT factor receptor-2 gene in Crouzon syndrome.";
 RL Hum. Mol. Genet. 4:1077-1082(1995).
 RN [11]
 RP VARIANT PS ALA-321.
 RX MEDLINE=95235551; PubMed=7719333;
 RA Lafenlle E., Wei M.H., Bonaventure J., Munnich A., le Merrer M.,
 RA Renier D.;
 RT "FGFR2 mutations in Pfeiffer syndrome.";
 RL Nat. Genet. 9:108-108(1995).
 RN [12]
 RP VARIANTS AS TRP-252 AND ARG-253.
 RX MEDLINE=95235562; PubMed=7719344;
 RA Wilkie A.O.M., Stanley S.F., Oldridge M., Poole M.D., Ashworth G.J.,
 RA Hockley A.D., Hayward R.D., David D.J., Pulleyn L.J., Rutland P.,
 RA Malcolm S., Winter R.M., Reardon W.;
 RT "Aper syndrome results from localized mutations of FGFR2 and is
 RT allelic with Crouzon syndrome.";
 RL Nat. Genet. 9:165-172(1995).
 RN [13]
 RP VARIANTS PS PRO-341; ARG-342 AND TYR-342.
 RX MEDLINE=95235563; PubMed=7719345;
 RA Rutland P., Pulleyn L.J., Reardon W., Baralster M., Hayward R.,
 RA Jones B., Malcolm S., Winter R.M., Oldridge M., Stanley S.F.,
 RA Poole M.D., Wilkie A.O.M.;
 RT "Identical mutations in the FGFR2 gene cause both Pfeiffer and
 RT Crouzon syndrome phenotypes.";
 RL Nat. Genet. 9:173-176(1995).
 RN [14]
 RP VARIANTS CS CYS/PS.
 RX MEDLINE=96203627; PubMed=8644708;
 RA Meyers G.A., Day D., Goldberg R., Daenli D.L., Przylepa K.A.,
 RA Abrams L.J., Graham J.M. Jr., Feingold M., Moeschler J.B.,
 RA Ramsley E., Scott A.F., Jabs E.W.;
 RT "FGFR2 exon IIIa and IIIC mutations in Crouzon, Jackson-Weiss, and
 RT Pfeiffer syndromes: evidence for missense changes, insertions, and a
 RT deletion due to alternative RNA splicing.";
 RL Am. J. Hum. Genet. 58:491-498(1996).
 RN [15]
 RP VARIANTS CS CYS-105; GLU-338; CYS-351 AND ARG-384.
 RX MEDLINE=97101656; PubMed=8946174;
 RA Pulleyn L.J., Reardon W., Wilkes D., Rutland P., Jones B.M.,
 RA Hayward R., Hall C.M., Brunton L., Chun N., Lammer E., Malcolm S.,
 RA Winter R.M.;
 RT "Spectrum of craniosynostosis phenotypes associated with novel
 RT mutations at the fibroblast growth factor receptor 2 locus.";
 RL Eur. J. Hum. Genet. 4:283-291(1996).
 RN [16]
 RP VARIANTS CS ILE-331; ASP-336--ALA-337 DUPL AND TRP-356--THR-358 DEL.
 RX MEDLINE=97114301; PubMed=8956050;
 RA Steinberger D., Mulliken J.B., Mueller U.;
 RT "Crouzon syndrome: previously unrecognized deletion, duplication, and
 RT point mutation within FGFR2 gene.";
 RL Hum. Mutat. 8:386-390(1996).
 RN [17]
 RP VARIANT PS CYS-290.
 RX MEDLINE=97295073; PubMed=9150725;
 RA Tartaglia M., Valeri S., Velardi F., di Rocco C., Battaglia P.A.;
 RT "TRP90Cys mutation in exon IIIa of the fibroblast growth factor
 RT receptor 2 (FGFR2) gene is associated with Pfeiffer syndrome.";
 RL Hum. Genet. 99:602-606(1997).
 RN [18]
 RP VARIANT CS L-252, VARIANT AS F-252, AND VARIANT PS F-252--S-253.
 RX MEDLINE=97156222; PubMed=9002682;
 RA Oldridge M., Lunt P.W., Zackai E.H., McDonald-Moggin D.M., Muenke M.,
 RA Woloney D.M., Twigg S.R.F., Heath J.K., Howard T.D., Hoganson G.,
 RA Gagnon D.M., Jabs E.W., Wilkie A.O.M.;
 RT "Genotype-phenotype correlation for nucleotide substitutions in the
 RT Ig11-Ig111 linker of FGFR2.";
 RL Hum. Mol. Genet. 6:137-143(1997).

```

RN [191]
RP VARIANT CS GTU-292.
RX MEDLINE=97297373; PubMed=9152842.
RA Steinberger D., Collmann H., Schmalenberger B., Mueller U.;
RT "A novel mutation (a886g) in exon 5 of FGFR2 in members of a family
RL with Crouzon phenotype and plagiocephaly."
RN J. Med. Genet. 34:420-422(1997).
[20]
RP VARIANTS CS VAL-276; CYS-301 AND SER-314.
RX MEDLINE=98180879; PubMed=9521581.
RA Steinberger D., Vriend G., Mulliken J.B., Mueller U.;
RT "The mutations in FGFR2-associated craniosynostoses are clustered in
RT five structural elements of immunoglobulin-like domain III of the
RT receptor."
RL Hum. Genet. 102:145-150(1998).
[21]
RP VARIANTS AS TRP-252 AND ARG-253.
RX MEDLINE=98112406; PubMed=9452027.
RA Tsai F.-J., Hwu W.-L., Lin S.-P., Chang J.-G., Wang T.-R., Tsai C.-H.;
RT Two common mutations 934C to G and 937C to G of fibroblast growth
RT factor receptor 2 (FGFR2) gene in Chinese patients with Apert
RT syndrome."
RL Hum. Mutat. Suppl. 1:S18-S19(1998).
[22]
RP VARIANT PS CYS-351.
RX MEDLINE=98356420; PubMed=9693549.
RA Mathijssen I.M., Vaandrager J.M., Hoogeboom A.J.,
RA Hesseljen Janssen A.L., van den Ouweland A.M.W.;
RT "Pfeiffer's syndrome resulting from an S351C mutation in the
RT fibroblast growth factor receptor-2 gene."
RL J. Craniofac. Surg. 9:207-209(1998).
CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH
CC FACTORS.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
CC TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DISEASE: DEFECTS IN FGFR2 ARE A CAUSE OF CROUON SYNDROME (CS),
CC ALSO CALLED CRANIOFACIAL DYSOSTOSIS TYPE I (CFDI). CHARACTERIZED
CC BY CRANIOSYNOSTOSIS (PREMATURE FUSION OF THE SKULL SUTURES),
CC HYPERTELORISM, EXOPHTHALMOS AND EXTERNAL STRABISMUS. PARROT-BEAKED
CC NOSE, SHORT UPPER LIP, HYPOPLASTIC MAXILLA, AND A RELATIVE
CC MANDIBULAR PROGNATHISM.
CC

Query Match 4.3%; Score 102.5; DB 1; Length 821;
Best Local Similarity 24.0%; Pred. No. 5.1;
Matches 78; Conservative 35; Mismatches 109; Indels 103; Gaps 18;

QY 182 CFSNFIASKL-----EAYEYAHISRLQADL--NCLSDL-----LKSSIQMKYIYN 225
DB 491 CGGYVYMAEAVGIDKDKPKEAVTVA-VKMLKDDATEKDLSDIVSEMKMKIGKHKIIN 549
QY 226 L---CGDFPLKSNFELYSELKLGANMLLETVPKPSKLEFTHHELRVPEYVYKLP 282
DB 550 LIGACTQDPL--YVIEYVASKGNLREYLARRPQME-----YSYINNVPEQMF- 600
QY 283 IFTNISKAPPHNIQIFGSAFV-----LSQAFKYIFNNSITVDFFAWMSKDTISP 334
DB 601 -----KDLVSCYOLARMEYLASQKCIHRDLAARNLVTENNMKIADJGLADINI 654
QY 335 DEHFATLIRVPGIGELISRSADVSDLSKTRLVKMNVEEGEFPYSCGSHLREVCYIG 394
DB 655 DYKKTITNRLP-----VKMAPELFLPRVYT--HQSVWMSFG 690
QY 395 -----AAELRWLIKQGHFANKFDSKVDILIKLCAEKLCEQORD-W 435
DB 691 VLMWEITLIGSPYGPVPEELFKLLKKGH-----MDKPPAN-----CTNE-LYMMARDCW 740
QY 436 IFLPSEK-LF-----MDRNLTTTS 453
DB 741 HAVPSGRPTFKQVLEDLDRILTLTT 765

```

```

FGFR2_MOUSE
ID FGFR2_MOUSE STANDARD: PRT; 821 AA.
AC P21803; 061342.
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FIBROBLAST GROWTH FACTOR RECEPTOR 2 PRECURSOR (FGFR-2)
DE (KERATINOCYTE GROWTH FACTOR RECEPTOR).
GN FGFR2 OR ECF1 OR BEK.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE=Brain, and Liver;
RX MEDLINE=92228773; PubMed=1373495;
RA Mansukhani A., Dell'Era P., Moscatelli D., Kornbluth S.,
RA Hanafusa H., Basilico C.;
RT "Characterization of the murine BEK fibroblast growth factor (FGF)
RT receptor: activation by three members of the Fgf family and
RT requirement for heparin."
RL Proc. Natl. Acad. Sci. U.S.A. 89:3305-3309(1992).
[2]
RN SEQUENCE FROM N.A. (SHORT FORM)
RX MEDLINE=91093977; PubMed=1846048;
RA Miki T., Fleming T.P., Bottaro D.P., Rubin J.S., Ron D.,
RA Aaronson S.A.;
RT "Expression cDNA cloning of the KGF receptor by creation of a
RT transforming autocline loop."
RL Science 251:72-75(1991).
[3]
RN SEQUENCE OF 477-821 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89219016; PubMed=2468999;
RA Kornbluth S., Paulson K.E., Hanafusa H.;
RT "Novel tyrosine kinase identified by phosphotyrosine antibody
RT screening of cDNA libraries."
RL Mol. Cell. Biol. 8:5541-5544(1988).
CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH FACTORS.
CC POSSESSES A HIGHER AFFINITY FOR ACIDIC THAN FOR BASIC FGF'S.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
CC TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
DB EMBL; M86441; AAA37286.1; -
DB EMBL; M63503; AAA39377.1; -
DB EMBL; M23362; AAA37285.1; -
DB PIR; A31378; TYMSBK.
DB PIR; A38429; A38429.
DB HSSP; P11362; IFG1.
DB MGD; MGI:95523; Fgfr2.
DB InterPro; IPR000719; Euk_Pkinase.
DB InterPro; IPR003006; Ig_MHC.
DB InterPro; IPR003598; Ig_C2.
DB InterPro; IPR001245; Tyr_Kin.
DB Pfam; PF00047; Ig_3.
DB Pfam; PF00069; Pkinase; 1.
DB PRINTS; PR00109; TYRKINASE.
DB SMART; SM00408; Igc2; 3.

```

RESULT 10

DR SMART: SM00219; TYRKC: 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP: 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TMR: 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DGM: 1.
 KW Receptor: Glycoprotein; Tyrosine-protein kinase; ATP-binding;
 KW Transferase; Phosphorylation; Transmembrane; Signal;
 KW Immunoglobulin domain; Alternative splicing.
 FT SIGNAL 1 21
 FT CHAIN 22 821 FIBROBLAST GROWTH FACTOR RECEPTOR 2.
 FT DOMAIN 22 377 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 378 398 POTENTIAL.
 FT DOMAIN 399 821 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 55 114 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 172 238 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 271 349 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 481 770 PROTEIN KINASE.
 FT NP_BIND 487 495 ATP (BY SIMILARITY).
 FT BINDING 517 517 ATP (BY SIMILARITY).
 FT ACT_SITE 626 626 BY SIMILARITY.
 FT MOD_RES 657 657 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT DISULFID 62 107 POTENTIAL.
 FT DISULFID 179 231 POTENTIAL.
 FT DISULFID 278 342 POTENTIAL.
 FT CARBOHYD 83 83 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 123 123 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 228 228 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 241 241 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 265 265 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 331 331 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPIC 37 37 E -> G (IN SHORT ISOFORM).
 FT VARSPIC 38 152 MISSING (IN SHORT ISOFORM).
 FT VARSPIC 152 152 MISSING (IN SHORT ISOFORM).
 FT VARSPIC 314 361 AGVNTDKIEILVIRNTPEDAGETGLAGNSIGISERS
 FT VARSPIC 314 361 AMTLVLP -> HSGISSNAVALNPTMAGEYICKVS
 FT VARSPIC 314 361 NYIGANOSAMTLVLPKQ (IN SHORT ISOFORM).
 FT CONFLICT 169 169 C -> Y (IN REF. 2).
 FT CONFLICT 187 187 S -> P (IN REF. 2).
 FT SEQUENCE 821 AA; 91983 MW; FCD28AD61F4414 CRC64;

Query Match 4.3%; Score 102.5; DB 1; Length 821;
 Best Local Similarity 24.0%; Pred. No. 5.1;
 Matches 78; Conservative 35; Mismatches 109; Indels 103; Gaps 18;

QY 182 CFSNIFLASKL-----EAVEYAHISRLQADL--NCLSDL-----LKSSIQWKYVIN 225
 DB 491 CFGGVVAEAVGIDKDKPKRAVYVA-VKMLKDATEKDLSDLVSEMEMKMKHKNIIIN 549
 QY 226 L--CGGDPLKSNFELVSELEKLGANMLETVKPPSKLERFTYHHELRVREYVKLP 282
 DB 550 ILGACITDDGL--YVIVEYASGNLREYLRARPPQME-----YSIDIRVVEQWTF- 600
 QY 283 IRTNISLEADPHNIQIFVGSAYFY-----LSQAFVKYIFNNSIVQDFPANSKDTYSP 334
 DB 601 -----KDLVSCYQLARGMEYLAQKCIHRDLARNAVLTENNVMKIAQGLARDINNI 654
 QY 335 DEHEWATLLVPGIPIGSISSACDVSDLOSKTRLVKWNYVEGFYPCSTSHLRVCITYG 394
 DB 655 DYIKATTNGRLP-----VKWMADEALDRVYT--HSDVWSFG 690
 QY 395 -----AAELRWLIKQHWFMANKFDSKDPILIKLAEKLEQORD-W 435
 DB 691 VLMEIFTLGGSPYPIGPIVELFKLKEGH-----RMDKPTN-----CTNE-LVMMRDOW 740
 QY 436 IILPSEK-LF-----WDRNLTITS 453
 DB 741 HAVPSQRTFKQVLEDLRLITLT 765
 RESULT 11
 DP03_UREPA

ID DP03_UREPA STANDARD; PRT; 1442 AA.
 AC 09PGB4:
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA POLYMERASE III POLC-TYPE (EC 2.7.7.7) (POLIII).
 GN POLC OR U0377.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Ureaplasma.
 OX NCBI_TaxID=134821;
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROVAR 3.
 RX MEDLINE=20500219; PubMed=11048724;
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
 RA Cassell G.H.;
 RT "The complete sequence of the mucosal pathogen Ureaplasma
 RT urealyticum."
 RL Nature 407:757-762(2000).
 CC -1- FUNCTION: REQUIRED FOR REPLICATIVE DNA SYNTHESIS. THIS DNA
 CC POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N-DEOXYNUCLEOSIDE TRIPHOSPHATE -
 CC N-PYROPHOSPHATE + DNA(N).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. POLC
 CC SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AE002135; AAP30786.1.
 CC InterPro: IPR000520; Exonuclease.
 CC InterPro: IPR003141; PHF_N.
 CC Pfam: PF00929; Exonuclease; 1.
 CC Pfam: PF02231; PHF_N; 1.
 DR SMART: SM00479; EXOIII; 1.
 DR SMART: SM00481; POLIIIa; 1.
 KW Transferase; DNA-directed DNA polymerase; DNA replication; Hydrolyase;
 KW Nuclease; Exonuclease; Complete proteome.
 FT DOMAIN 409 576 EXONUCLEASE.
 FT SEQUENCE 1442 AA; 166227 MW; 834C3EF722ACA6D1 CRC64;

Query Match 4.2%; Score 101.5; DB 1; Length 1442;
 Best Local Similarity 18.9%; Pred. No. 13;
 Matches 100; Conservative 88; Mismatches 181; Indels 161; Gaps 27;

QY 33 NVRRLEPQ-----KDIIVEYSLSTS-----PVRNRYTHVDE-----VRYEVN 72
 DB 361 NVYQAVPEIAKSKKODLKI--YGLETEHELDQIPLVLA---VRQMLDNATYVIFDIE 415
 QY 73 CGGIYE-----QEPLTIGKSLTERRRDIIDLEDQDVYAMSDCDIYITLNGY 119
 DB 416 TTGLPFPNDEILIEFGAVIMQNNKQIGEKIOFIKPIQOINEN---VTNLNISQEMWN 471
 QY 120 A--QKIVSEKESFPLAYSLVHK-----DAIWERLII 151
 DB 472 AIDKEIALKIKEIFDDHILVAHNGINPDININORLLKWLGEPLKNSIDTLMISRAIN 531
 QY 152 AI-----YQHNITCIHYDR-KAPDTFKYAMNNLAKCF--SNIF-IASKL 192
 DB 532 PKRSRLGAICKKRYEVDYDESAHRAVDIAVLAVFYKMKNNLENDGIIINLSINKRL 591
 QY 193 EAVEYAHIS-----RLQADLNCISDLTKSSIQWKYVINLGGQDPLKSNFELVSEL 243
 DB 592 QTTMLKNRBFQGWINILYIKQANVAKMDYELVLSISHTDMYTR-----PTITTSFLANK 645

```

OY 244 KKLNGANLMTVKKPNSKLERFTY---HHELRVY--PREYVKLIRTN-----ISKAPP 293
DB 646 DKLIIISNIH-----ESLINALSKNDEEIKRLIQRYDITLPSLOSQHLYVAKITTI 700
OY 294 HNIOFVGSAAVFLSQAVKFIYFNNSIVODEFWANSKDTY---SPDEHFMATLIRVPGIPG 350
DB 701 ENVQ-----KAFKRLIYALALEINKITIIYSSSPYEFKDKKFEVDVYVNTKGLEG 749
OY 361 EISSAOD--VSDIO---SKTRLKMNMYEGFFPSCGTGSLRBCVCIYGAELMLKDG 405
DB 750 KAHFANAVYVPDLEIYDQKNALDELAYLE-----DEKLILILINEN 791
OY 406 HMFANK-FDSKVPILIKCLAEKLE---EQQRWITLPSKELFMDRLNLT 451
DB 792 PVHINSWEDSDIOPLKEGLYAPKKEGVQDKTIDYVHTAKIY-GENLPT 840

RESULT 12
YCF2_PINTH STANDARD; PRT; 2054 AA.
AC P41653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL 244.6 KDA PROTEIN YCF2 (ORF 2054).
GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugita M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii."
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -1- FUNCTION: NOT YET KNOWN.
CC -1- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D17510; BAA04460.1; -
DR InterPro: IPR001939; AAA_subfam.
DR Pfam: PF00004; AAA_1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 24460 MW; 6F5E92D07B33A9A CRC64;

Query Match 4.2%; Score 101.5; DB 1; Length 2054;
Best Local Similarity 21.5%; Pred. No. 21;
Matches 72; Conservative 60; Mismatches 98; Indels 105; Gaps 20;
OY 15 KVFILFLTLV-LLSLKLNLNRLRFPKQDIYLVY-SLSTSPFYRNRVTHVKDEVRVYN 72
DB 864 QTFVILIDLYRSLNLRNLNLP--FVREKRLSIIEISITPLRKEOI-----VNFEN 914
OY 73 -CSGIYEOEPLIGKSLERRDIIIDEDDVAMTSCDQIYCLRGYAKLVS-----K 126
DB 915 FCQPFKRSDEENNFCQCFR-----GFSSNGLIQT-RSYQDDILSEMSNK 962
OY 127 EKSAPFIYSLVHKDAIYMERLHAIIYNCHNIYCIHYDRKAPDTFVANNLAKCFST 186
DB 963 NEEIFPRF-----QDWFEVTECLNKNKIYNE-----DIDGRS-----TISNSKEEONI 1004

```

```

OY 187 FIKSKLEAVYAHISRLQADLCLSLKSSIQW-----KYVINLCQGDPLKSNEL 239
DB 1005 YRISQIDSI-----FSKW-----DLFTYMPWFTSAWCXYIENML-----LDT 1043
OY 240 VSELKRLKLAN-----MLETFKPPNSKLERFTYHHELRVYVVKLPIRTNISKAPPN 295
DB 1044 LSLT-LLHGSNPFVSLQINIK-HNLLKR-NLWELSHLMPDIOCKLRTNLI----- 1094
OY 296 IQFVGSAAVFLSQAVKFIYFNNSIVODEFWANSKD 330
DB 1095 -----KEFFPSNNEKDFPPYCKD 1112

RESULT 13
YCF2_TOBAC STANDARD; PRT; 2280 AA.
AC P09976; P09977;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL 267 KDA PROTEIN YCF2 (ORF 2280).
GN YCF2.
OS Nicotiana tabacum (Common tobacco).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. BRIGHT YELLOW 4;
RA Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N.,
RA Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,
RA Yamaguchi-Shinozaki K., Ohno C., Torazawa K., Meng B.Y., Sugita M.,
RA Denc H., Kamagashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
RA Tohoh N., Shimada H., Sugita M.;
RT "The complete nucleotide sequence of the tobacco chloroplast genome:
RT its gene organization and expression."
RL EMBO J. 5:2043-2049(1986).
RN [2]
RP REVISIONS.
RA Sugita M.;
RT Submitted (Feb-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: NOT YET KNOWN.
CC -1- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z00044; CAA77427.1; -
DR EMBL: Z00044; CAA77438.1; -
DR PIR: A05204; A05204.
DR PIR: A05205; A05205.
DR InterPro: IPR001939; AAA_subfam.
DR Pfam: PF00004; AAA_1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2280 AA; 266812 MW; E246D5F3D902C06D CRC64;

Query Match 4.2%; Score 101.5; DB 1; Length 2280;
Best Local Similarity 20.0%; Pred. No. 24;
Matches 88; Conservative 67; Mismatches 169; Indels 117; Gaps 20;
OY 9 KATLQGVILFLTLVLLSLKLNLNRLRFPKQDIYLVYSLSTSPFYRNRVTHVKDEVR 68
DB 761 KDTLNHRTIMKYTINQVLSNLKSKCKKMPEDILISRTFSRSMNDP-----DAYR 810

```


Job time: 205 sec

```

RESULT 15
VOO1_VACCC STANDARD; PRT; 666 AA.
AC P21093;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN O1.
GN O1.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; Pubmed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus.'";
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M35027; AAA48053.1;
CC PIR: D42510; D42510.
SQ SEQUENCE 666 AA; 77577 MW; 46DF329DDEB66EF CRC64;

Query Match 4.2%; Score 99.5; DB 1; Length 666;
Best Local Similarity 22.0%; Pred. No. 6.4;
Matches 75; Conservative 50; Mismatches 111; Indels 105; Gaps 18;

QY 46 VEYSLSSTFFVNRRTYHVDKRYEVNCSGIYEPLLEIGKSLER----- 92
DB 292 VEVNLSIYDE--RYRQFADEFRDYIM--IKERQITWQSGDRIRFRPRMSLRSTIRK 345
QY 93 RDIIIDLED-----DDVY-----AMTSPCDIYQTLRGYAQ 121
DB 346 KDTDSLEDILAHIDNARKMSKVISIEDVERIISFRINPCVVRKRLSDIDIKTKI--MYL 403
QY 122 KLYSKEEKSFPYASLVYHKDAIMVERLIHAYNOHNICYIHYDRK-----ADTFKVA 175
DB 404 KIV-KDWSKSCALTLISAI--KGIMVTDIINTVLSK---ILHHRNVEFYVLTSVENKEIA 455
QY 176 MNNLAKCFNSNIFIAKSLKLEAVEYAHISRLQADLNC-----LSDELKSSIQMKYVINLCG 228
DB 456 VCNCSRCIS-----LEFRELKSVRCDLRTDDGLDRLYDLTRALHGNKINQNLIG 505
QY 228 QDF--PLKSNFELVSELKKNLGNANLETVKPPNSKLEFRFYHHLELRVPEYVXLPRTN 286
DB 506 QRCWGPLETEMLFENKKRKLN--NLMEYIKIS---DMLVYGHSI-----EKTLPIDTS 554
QY 287 ISKEAPPHNICIFVGSAYFVLSQAFVK-YIFNNSIVODEFA 326
DB 555 LSFKLSDVTMS-----VLNDQYAKIVIFPNTIIEYIIA 587

```


Fri May 3 10:57:56 2002

us-09-645-192-2.rsp

Page 13

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2002, 07:43:09 ; Search time 19.44 Seconds

(without alignments)
1775.055 Million cell updates/sec

Title: US-09-645-192-2

Perfect score: 2389

Sequence: 1 MKIFRCYKHTLQCKVILF.....DWITLPSEKLFMDENLTTS 453

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1:
2: p1r2:
3: p1r3:
4: p1r4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	863.5	36.1	428	2 A46293	beta-1,3-galactosyl
2	711	29.8	400	2 A46297	beta-1,6-N-acetylgl
3	326	13.6	445	2 T21262	hypothetical prote
4	323	13.5	425	2 T20207	hypothetical prote
5	315.5	13.2	454	2 T21261	hypothetical prote
6	296	12.4	401	2 T24929	hypothetical prote
7	294	12.3	402	2 T24928	hypothetical prote
8	275	11.5	402	2 T24930	hypothetical prote
9	245	10.3	459	2 T24742	hypothetical prote
10	229.5	9.6	454	2 T25382	hypothetical prote
11	220.5	9.2	489	2 T24013	hypothetical prote
12	219.5	9.2	470	2 T32137	hypothetical prote
13	219.5	8.5	486	2 T21566	hypothetical prote
14	203.5	8.5	472	2 T33384	hypothetical prote
15	202.5	8.5	753	2 T24745	hypothetical prote
16	197.5	8.3	448	2 D85042	probable glycosyla
17	193.5	8.1	478	2 T27714	hypothetical prote
18	188	7.9	467	2 T21408	hypothetical prote
19	187	7.8	447	2 T00906	hypothetical prote
20	182.5	7.6	513	2 T20068	hypothetical prote
21	178	7.5	367	2 T28892	hypothetical prote
22	175.5	7.3	406	2 F96571	hypothetical prote
23	168	7.0	395	2 B96735	unknown protein f2
24	166.5	7.0	434	2 T51450	hypothetical prote
25	132	5.5	630	2 T02524	probable RING zinc
26	131	5.5	384	2 T08940	hypothetical prote
27	126	5.3	371	2 T21819	hypothetical prote
28	116	4.9	895	2 E64431	UDP-glucose 6-dehyd
29	108.5	4.5	661	2 S50734	WH13 protein - yea

30	108	4.5	702	2 E69498	hypothetical prote
31	107	4.5	717	2 S78177	hypothetical prote
32	107	4.5	1072	2 T37742	serine threonine-p
33	107	4.5	2265	2 T26183	hypothetical prote
34	106.5	4.5	1347	2 T41321	B7B domain and Ank
35	104	4.4	364	2 T48628	hypothetical prote
36	104	4.4	2485	1 H71621	serine/threonine-s
37	104	4.4	2628	2 S59413	probable membrane
38	103.5	4.3	588	2 H82877	hypothetical prote
39	102.5	4.3	415	2 T55233	heparin-binding fi
40	102.5	4.3	707	2 A54846	fibroblast growth
41	102.5	4.3	707	2 A38429	keratinocyte growt
42	102.5	4.3	769	2 S16236	fibroblast growth
43	102.5	4.3	820	2 S17295	fibroblast growth
44	102.5	4.3	821	1 TVMSBK	fibroblast growth
45	102.5	4.3	821	1 TVHDS2	fibroblast growth

ALIGNMENTS

RESULT 1
A46293
beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase
C:Species: Homo sapiens (man)
C>Date: 20-Oct-1993 #sequence-revision 18-Nov-1994 #text-change 21-Jul-2000
C:Accession: A46293
R:Bierhuizen, M.F., Fukuda, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 9326-9330, 1992
A>Title: Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcN
A:Reference number: A46293; MIMD:93028457
A:Accession: A46293
A:Status: preliminary
A:Molecule type: RNA
A:Residues: 1-428 <BI>
A:Cross-references: GB:M97347; NID:G183440; PID:AAA35919.1; PID:G183441
A:Experimental source: Hc-60 cells
A>Note: sequence extracted from NCBI backbone (NCBI:115900, NCBI:P:115901)
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match	36.1%; Score 863.5; DB 2; Length 428;
Best Local Similarity	No. 43.1%; Pred. No. 6.2e-56;
Matches	187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;
QY	28 LKLLVRRLEF--PQDILV- EYSLSTSPFYRNRTYKDE---VRY-----EV 71
DB	1 MLRTLLRRRLFSYPTKYFVWLVLSLTFPSVLR---IHQKPEFVSRYHLELAGENPSSDI 57
QY	72 NCSGIYOEPLFEGK-----SLEIRRDIIIDEDDDVVAAMSDCDIYQTLRGVAKLVS 125
DB	58 NCRTKVIQGVNLEQKVKLELTVKFKRP--KWPDPDIYINMSDCSSFKRKRIYVEPLS 115
QY	126 KEKSPPIASLVYHKDAIWEVERLHAIYNOHNYICIHDRAPDFTKYAAMNNLAKCFSN 185
DB	116 KEAEAPPIASVYIYVHHKIMDLRLAIYMPNFCVYHDTSEDSYLAAVGVASCSFN 175
QY	186 IFASLLEAVEVAHISRILOADLNCISDLKSSIQWYVNLGQDPELKSNEPELSEKK 245
DB	176 VFASRLSESVYASMSRVADLNCMDLYAMSNKKYILNLGMPFITNIEIYRKLL 235
QY	246 LNCANMLEVTPKPNLSKLEFTYHHELRVPEYV--KLPFRINISKAPENHIQIFVGS 303
DB	236 LMGENNLETERFMSHKEEMV-----KKRYEVNGLT-TNGTYAMLPLETPFSGSA 287
QY	304 YFTLSQAFYKTIINNSIYVDFEAFMSKDTSPDEHFAFLIYVPGIGETISRSNO-DVSD 362
DB	288 YFVVSSEIYGVYLVQNEKIKQIMEMADDTSPDEILMATIQRLPEVPGSLPAHKKIDLSM 347
QY	363 QSKTRIVKNNYEGEF-----YPSCTGSHLRSCVIGAAELRMILKDGHPANKEDSKYD 417
DB	348 QAVARFVKNQYFEGDVSKGAPYRPPDGVVHVSVCIGAGADLMMMLKKNHLFANKFQVDV 407

OY 418 PILIKCLAEKLEEQ 431
 Db 408 LEAIOCLDEHLRHK 421

RESULT 2

A46297
 beta-1,6-N-acetylglucosaminyltransferase (EC 2.4.1.-) - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C:Accession: A46297
 R:Bierhuizen, M.F.; Mattei, M.G.; Fukuda, M.
 Genes Dev. 7, 468-478, 1993
 A:Title: Expression of the developmental I antigen by a cloned human cDNA encoding a mem
 A:Reference number: A46297; MUID:93194065
 A:Accession: A46297
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-400 <B1E>
 A:Cross-references: GB:219550; NID:g296531; PIDN:CAA79610.1; PID:g296532
 A:Note: sequence extracted from NCBI backbone (NCBIN:127166, NCBIIP:127167)
 C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 29.8%; Score 711; DB 2; Length 400;
 Best Local Similarity 43.7%; Pred. No. 9e+45;
 Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;

OY 108 SDCDDYQTLRGYAKLVKSEKSPFIAYSLVVKDAIWERLHAIYNQHIYCIHYDRK 167
 Db 70 SSCCKEYLQSHYITAPLEKSEADFLATVIMHHFDFALFRATYVQNIYCVHDEK 129
 OY 168 ADPTKVMNNLAKCFNFIASKLEAVEYAHISRLQDLNCLSKSIQKVIYINLC 227
 Db 130 ATTERKDAEQQLSCFPNAPFLAKMEPVYGISRLQDLNCLSKSAFEVSMKYVINTC 189
 OY 228 GODPFLKSNFELVSLKLNKANMLETVKPNKSLERFTY-HNELRVYEVYKPIRTN 286
 Db 190 GDDFLKTNKEIYQYLRKCKNINIPGVLPFAHAIGRKRYHQHLEKELSYV--IRTT 246
 OY 287 ISEKAPPHNIOIFVGSATFVLSQAVKTIFFNNSTIYQDFNASKDTYSDEHFWATLIVP 346
 Db 247 ALKPPPHALITTFGSAYVAASREPANVLHDPRRAVDLQMSKDFSPDEHFWYTLNIP 306
 OY 347 GIPGEISRAQDVSDLSQSTRLVKWNYYEGEFPYFCTGSHRSVCIYGAELRMILKDG 406
 Db 307 GYFGSMNPNS-----WTGNLRAIKMSDMED-RHGCGHHYHGICTIYNGDLKMLVNSPS 360
 OY 407 WFANKFDSKVPILIKCLAEKLEEQ 433
 Db 361 LFANKFELNTYPLTVECL--ELRHRR 385

RESULT 3

T21262
 hypothetical protein F22D6.11 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T21262
 R:Wilkinson, J.
 submitted to the EMBL Data Library, April 1996
 A:Reference number: Z19397
 A:Accession: T21262
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-445 <WIL>
 A:Cross-references: EMBL:Z71262; PIDN:CAA95817.1; GSPDB:GN00019; CESP:F22D6.11
 A:Experimental source: clone F22D6
 C:Genetics:
 A:Gene: CESP:F22D6.11
 A:Map position: 1
 A:introns: 65/3; 136/3; 180/3; 256/3; 281/3; 403/3

Query Match 13.6%; Score 326; DB 2; Length 445;

Best Local Similarity 25.1%; Pred. No. 2e+16; Matches 108; Conservative 85; Mismatches 176; Indels 62; Gaps 15;

OY 25 LLSLKLNVRLFPQKDIYVEYSLSSTSPVARNRYHAKDEV--RYEVNCGIY---E 78
 Db 17 LVFLFLFYKPLVPR---ILSLSNRRNPQETSILSKLENDLDDLDINCININGSKN 73
 OY 79 QEBLEIGKSLERIRRIIDLEDDVYAMSDDDIYQTLNGYAKLVKSEKSPFIAYSLV 138
 Db 74 RNLRLVNSRST-----EKLSTDRQCTLKSLFRNKVPLSPDESEFPLSTGL 124
 OY 139 VHKD--AIWER---LIHAIYNQHIYCIHYDRKAPDFKVMNNLAKCFNFIASKLE 193
 Db 125 VYKELSOVTLQVLFMLSSIYHPQNECYIAGVNSAPIRQNLKELSNCFNSIHF-MKRP 183
 OY 194 AVEYAHISRLQADLNLCLSKSIQKVIYINLCGDDPFLKSNFELVSLKLNKANMLE 253
 Db 184 PIDWGSHEIINSAYDCLFELSHLKSDMRYFYQISGVDIPLKTNLEWQILKHLNG---- 238
 OY 254 TVKPPNKLERFYHHELRRVPEYVYKLPFRTNISKAPPHNIOIFVGSAYFVLSQAEVK 313
 Db 239 -----TANVEIK--PYQYQRL--RQKNETQSP--LPFKSSLISLIPREAN 279
 OY 314 YIFNNSIVQDFNASKDTYSPDEHFWATL---RVPGIGELISRAQDVSDLSQSTRIV 369
 Db 280 HLSSSSIPQDLFLRLNTGJADEGFWGTLEGNKNLFDIGSLNFKEMISYKNNVETNLT 339
 OY 370 ---KWNYY---EGFEYSCGTGSHLSRVCIYGAELRMILKDGHWANKFDSKVPILIK 422
 Db 340 PTDGWRVYIRDDIWSKPNHNMKAGSCVFGIGDVPRLKSAALVAHNFYKSEBEAYF 399
 OY 423 CLAEKLEEQ 433
 Db 400 CL---LKEHRR 407

RESULT 4

T20207
 hypothetical protein C54C8.11 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T20207; T27478
 R:DeBoson, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19237
 A:Accession: T20207
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-425 <WIL>
 A:Cross-references: EMBL:Z83102; PIDN:CAB05469.1; GSPDB:GN00019; CESP:C54C8.11
 A:Experimental source: clone C54C8
 R:McLay, K.
 submitted to the EMBL Data Library, October 1998
 A:Reference number: Z20372
 A:Accession: T27478
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-425 <M12>
 A:Cross-references: EMBL:AL032664; PIDN:CAA21769.1; GSPDB:GN00019; CESP:C54C8.11
 A:Experimental source: clone Y91F4A
 C:Genetics:
 A:Gene: CESP:C54C8.11
 A:Map position: 1
 A:introns: 29/2; 117/3; 161/3; 195/3; 237/3; 266/3; 295/1; 313/3; 355/1; 374/2

Query Match 13.5%; Score 323; DB 2; Length 425;

Best Local Similarity 22.6%; Pred. No. 3.1e+16; Matches 102; Conservative 85; Mismatches 176; Indels 88; Gaps 14;

OY 16 VFILFLTWLSSLKLNVRLFPQKDIYVEYSLSSTSPVARNRYHAKDEVRYEVNCS 74

```

Db      5  IFGLFLLIFVTPLEPLVYLIVLEIFESQNPASTIDIFS-----TSQCDIYQILRGYAOXLVSK 126
QY      75  GITYOELELEIGKSLERIRDIIDLEDDVYAM-----TSQCDIYQILRGYAOXLVSK 126
Db      45  KMPDGDVSLA-----RGALFEFDEDELLAQILKLSGQEMACAEFOKIFGFEQPTSQ 98
QY      127  EKSFPPIAYSLVYHKDAIWEERLIHAAYNOHNIYCIHYDRKADTFKVMANNLAKCFSN 186
Db      99  EELEFPPIAYGMLVHGDVQJLSLLSATYQPOQCLAVDGNSSVEFIQVRLMSRCYGN 158
QY      187  --FIASLEAVEYAHISRLQADNCLSDLKSSIQWKYVINLCGQEPILKSNEFLVSEK 244
Db      159  QYFTTDEIRRCQGEILTSV---FOCVDYLLAKLPSPDMKYYFOYLSGVADPLKSNLEMTIRIK 215
QY      245  KLVGANMLETVKPPNSKLEERTYHHELRVRYEVYKLPITNTISKAPRNQIIFGSAV 304
Db      216  ALNG-----SERAEI--LPFEFYRLNKRKRPMSPLPYKTSL---SAT 253
QY      305  FVLSQAFVKYIFNNISIVQDFPAMSKDTPSPDEHFWATLIRVPGI-----PGEISRSA 356
Db      254  F--SRKSANFMVNSEKYLEQIDFLRGTTCADESIMATINAGNPKVIRFSELPMPGDPDAX 311
QY      357  QDVSDL-QSKTRLYKM-----NYYEGFFFP-----SCGSHLRSCVIGAEELRM 400
Db      312  WIKHNYRIRIGKLGKYEENQKIDNGYVSRIOQYVNPAPVKCKGYIRLSGCVGYDLPN 371
QY      401  LKDGHWFAFKPDSKYDPIILKCLAEKLEEQ 431
Db      372  LINRHELVAHKLYFSYQPAFMCLVENSROK 402

RESULT  5
T21261
hypothetical protein F22D6.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #extl_change 15-Oct-1999
C:Accession: T21261
R:Milkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19397
A:Accession: T21261
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-454 <WILL>
A:Cross-references: EMBL:Z19397; PIDN:CA95816.1; GSPDB:GN00019; CESP:F22D6.12
A:Experimental source: clone F22D6
C:Genetics:
A:Gene: CESP:F22D6.12
A:Map position: 1
A:Introns: 62/3; 92/2; 190/3; 266/3; 291/3; 412/3

Query Match      13.2%; Score 315.5; DB 2; Length 454;
Best Local Similarity 24.3%; Pred. No. 1.2e-15;
Matches 111; Conservative 79; Mismatches 190; Indels 77; Gaps 16;

QY      8  FKHTLOQKVFILFLTLMILSLKILNVRRLFPQKDIYLVYSTSPFVNRRTYHKDEV 67
Db      6  FKSLTIAIFLFI-----YFVESLFPKQ---EDKNVSKQFLKSITTTASDSY 52
QY      68  ---RYEVNGSIGIYOELEIGKSLERIRDIIDLEDDVYAMSDCDIYQTLRGYAOXLV 124
Db      53  LLDNMEINCSNII--KQYKTEKRLIDIMHDI---EEOLESCTNKQCLTKLTERFVTNPM 107
QY      125  SKEKSPPIAYSLVYHKDA-----IWEERLIHAAYNOHNIYCIHYDR 166
Db      108  SAEKHEPPIYGLMVLVYDLPVYPPARMEFLKHLLEINLOVLPILSSITVHQNQYCLAVGE 167
QY      167  KAPDTEKVMANNLAKCFSNIFIASKLEAVEYAHISRLQADNCLSDLKSSIQWKYVINL 226
Db      168  NSAPIFONLREYVSTCFSNVHF--MKRPPISMGSHIIDSYYDCLLEFLSHLETWAKRYOYL 226

```

```

QY      227  CGQDPILKSNEFLVSEKILKLVGANMLETVKPPNSKLEERTYHHELRVRYEVYKLPITRN 286
Db      227  SGVDIPKTNLEMLVOLKILNLTSTNVEIT--NYQARLTGKNE----- 267
QY      287  ISKEAPPNIIQIFVGSAYFVLISQAFVKYIFNNISIVQDFPAMSKDTPSPDEHFWATLIRV 346
Db      268  --NESP---LPFKSSLIAIIPRKANOLASNTARKLLEFLMNTIEDGFWGTLPFGNK 322
QY      347  ---GIPGEL-SRSADVDLSQK---TRLVKNMY---EGFTFPSTGSHLR-SVQIYGA 396
Db      323  DQFNISGSIKSDMEYRNQNNIFNPPTDGSYIYSRQIWDPELCKYKMDSDCVFQIG 382
QY      397  ELRWIKDGHWFANKPDSKVPDILIKCLAEKLEEQ 433
Db      383  DVPRLTISKALVAHKFYLKSEPEAVFCL--LKEHR 416

RESULT  6
T24929
hypothetical protein T15D6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #extl_change 15-Oct-1999
C:Accession: T24929
R:Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19356
A:Accession: T24929
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-401 <WILL>
A:Cross-references: EMBL:Z283125; PIDN:CAB05620.1; GSPDB:GN00019; CESP:T15D6.2
A:Experimental source: clone T15D6
C:Genetics:
A:Gene: CESP:T15D6.2
A:Map position: 1
A:Introns: 45/3; 75/2; 111/3; 151/3; 227/3; 252/3; 296/2; 359/3

Query Match      12.4%; Score 296; DB 2; Length 401;
Best Local Similarity 23.4%; Pred. No. 2.8e-14;
Matches 86; Conservative 68; Mismatches 157; Indels 56; Gaps 7;

QY      61  THAKDEVREYVNGSIGIYOELEIGKSLERIRDIIDLEDDVYAMTSCDIYQTLRGYA 120
Db      34  TOLLDDL--EINCTNII--QGLAKNEKOLEINTKMI---EDKLNSTDRCHTVKSMFRFN 86
QY      121  QKLVSEKESFPPIAYSLVYHKDAIWEERLIHAAYNOHNIYCIHYDRKADTFKVMANNLA 180
Db      87  EVPLSEEFARFPLSYGLLYVKELSGVLFMLSSIYQPOQNEYCIAVGENSASTFLILLELS 146
QY      181  KCFNSIFIASKLEAVEYAHISRLQADNCLSDLKSSIQWKYVINLCGQDPILKSNEFLV 240
Db      147  DCFPN-----KRPFTWGSYEIINSYVYDCLKFLSKKNMKYFOYLSGVDPILKTNLEMV 201
QY      241  SELKLVGANMLETVKPPNSKLEERTYHHELRVRYEVYKLPITNTISKAPRNIIQIFV 300
Db      202  RIKKSLINGANVEIKYENRRL-----LQONETESPPLPFK 237
QY      301  GSAYFVLSQAFVKYIFNNISIVQDFPAMSKDTPSPDEHFWATLIRVPGI---IPGEISRSAQ 357
Db      238  SSLSLIPRKANYVLASSSIPOQLLEFLRNTVWADGEGWGLTFNGKGLFDPVGSIN---- 293
QY      358  DVSDDLQSKTRLYKMYVYEGFFPSCGSHLRSCVIGAEELRWIKDGHWFANKPDSKVD 417
Db      294  -----FEHQIWFESGCHNMKMGSCVFGIGDVSNNLQAKALVAHKLYLISE 340
QY      418  PILIKCL 424
Db      341  PEAYFCL 347

RESULT  7
T22188

```


Db 73 KIEHFYSKYRRESETAHVDCGRILSGDKYLQTVSG--ENRIPMVENPNPDKMSCSAVMD 130
 QY 133 -----IAYSLVYHDAIMVERLIIAIIYNQNIYCIHIDRKAPDTFKYAMN 177
 Db 131 RIIFHDHILRPLKGVAFARLVYDYDVEQVMQSYHPQSFCAIDKAPFRFKQMR 180
 QY 178 NLACFSNIFASKLEAVEVAHISRLQADINCLSDLKSSIQWKYVNLGODEPLKSNF 237
 Db 191 AMAACLPVLLLPDQEPIDSGNHVNLALHNCCLALLNKP-GWNYAALLQNHDLITTSYV 249
 QY 238 ELVSELKLNANMLEYKPPNSKLERFTYHEHLRVPYEVKLPITRNISKEAPRNIIQ 297
 Db 250 ELEQVTEWLGAGANDVELL-PEAQRLDEENFKWD---PSLIKMPDESKYDETLLNEKIK 304
 QY 298 IFVGSAYFLVLSQAFVKYIFNNSIVQDFP-AMSKDTYSPDEHFATL--IRVPGIGETSR 354
 Db 305 FSKGVSQSGMSRAAVDMWTRKVNISTYIDQNGRMGVDEMLISSLSIFLGMFGHF-- 362
 QY 355 SAQVSDLOQSTRLVKNYIEGFYPSCTGSHLR-SVCIYGAELEWLLIDGHWFKAFD 413
 Db 363 TDQCLEKEKEPFPMVQWMEDESY--CASKYKRHNVCIIIGIEYLRVAVASFPTLMFKML 419
 QY 414 SKVDPILIKCLAEKLEBO---ORDWITLPEKLFMDRNLTTS 453
 Db 420 PPFNSITECTAELLNRTFMGQNDH---PLEEERY-KMYTVS 459

RESULT 10
 T25382
 hypothetical protein T27F6.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T25382
 R:Doonan, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: 220026
 A:Accession: T25382
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1454 <WTL>
 A:Cross-references: EMBL:Z82060; PIDN:CA804881.1; GSPDB:GN00019; CESP:T27F6.1
 A:Experimental source: clone T27F6
 C:Gene: CESP:T27F6.1
 A:Map position: 1
 A:Introns: 80/3; 191/1; 278/1; 325/3; 365/2; 407/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 9.6%; Score 229.5; DB 2; Length 454;
 Best Local Similarity 22.7%; Pred. No. 2.5e-09;
 Matches 103; Conservative 68; Mismatches 202; Indels 81; Gaps 19;

QY 17 FILFLTLMLSLKLLNVRLLFPQKDIYLVESLSTSPFYARNRYTHVKDEVREVCNGSI 76
 Db 10 FIL-----LVALICTLN-----WFLYNYQV-----YRNHWSTYTONSSFEQDIKFE 50
 QY 77 YEQEPLEIGKSLERRRDIIDLEDDV---VAMTSDCDIYQTLRG-AAOKLVSKSEKSPF 132
 Db 51 YPTSNKDV---FYVRRRP---ETENNCGOVLAGDIAIKTYVGEIRIKIAENESLIMS 103
 QY 133 -----IAYSLVYHDAIMVERLIIAIIYNQNIYCIHIDRKAPDT 171
 Db 104 CEAVMDRLISRDHVLRLPLGVAFARLVYDYDVELEIKHVEKSYHPQSFCAIDKRAKE 163
 QY 172 FKVAMNNLAKGFSNIFI-----ASKLEAVEVAHISRLQADINCLSDLKSSIQWKY 222
 Db 164 FKERNQAAVASCPLPVLLPGRFFKNPFIHDSVDSHGHTNL-AHYNCLRALINKP-GWNY 221
 QY 223 VINLGGODPLKSNFELVSELKLNANMLETVKPPNSKLERFTYHEHLRVPYEVKLP 282
 Db 222 AILLQNHDLITKSYVELEKIFNWLGGANDV-AIRPELGRLDX-----KHFKWDPMSLK 273

QY 283 IRTNISKEAP---PHNIQIFVGSAYFLVLSQAFVKYIFNNSIVQDFP-AMSKDTYSPDEHF 338
 Db 274 LFRNESEIDPVIINTTLKFAKGAVOSSLSRAAVDMWTRFVDTLTITDQNNHGTIVQDQF 333
 QY 339 WAT-LIRVPGIGETSR-BSAQVSDLOQSTRLVKNYIEGFYPSCTGSHLR-SVCIYG 394
 Db 334 TQAFQSDLEFGMGHFTDKCIKKGITTEGTRFAQCTHD---QSKCASKSRHGICING 390
 QY 395 AALRLRLIDGHWFKAFKPSKVDPIILIKLAEL 428
 Db 391 IEHLSMKARSEHLMFNKVLPLFDYSITECTAELL 424

RESULT 11
 T24013
 hypothetical protein R07B7.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24013
 R:Harris, B.
 submitted to the EMBL Data Library, July 1996
 A:Reference number: 219830
 A:Accession: T24013
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1489 <WIL>
 A:Cross-references: EMBL:Z75955; PIDN:CA800115.1; GSPDB:GN00023; CESP:R07B7.6
 A:Experimental source: clone R07B7
 C:Gene: CESP:R07B7.6
 A:Map position: 5
 A:Introns: 42/3; 80/2; 235/3; 287/1; 372/2; 416/1

Query Match 9.2%; Score 220.5; DB 2; Length 489;
 Best Local Similarity 22.0%; Pred. No. 1.3e-08;
 Matches 102; Conservative 77; Mismatches 209; Indels 75; Gaps 19;

QY 22 TLWLISLKLNLNVRLLFP---QKDIYLVESL---STSPFYARNRYTH-----LEDDVYVAMTSDC 62
 Db 16 TLIKLEMSATRYKRLTIDIDLADINTARVSLQDNGKNSPIILRTSQEKPRKRNTIKWMD 75
 QY 63 ---VKDEVREYVCSGIYQEPLEIGKSLERRRDIID-----LEDDVYVAMTSDC 110
 Db 76 NYTTADESQF-LNCSEMIK-----NNKDYIELYVNNGRKLDNERLFEILPMDC 122
 QY 111 -DIYQTLRGYAOKLVSEKESPFIAVSLVYHDAIMVERLIIAIIYNQNIYCIHIDRKAP 169
 Db 123 PSIKNRIYG--DMPSPFLPKRIAFVRTIYKIELQELALLSYHPDNYVCFVMDSKST 179
 QY 170 DTFKVMNNLAKGFSNIFIASKLEAVEVAHISRLQADINCLSDLKSSIQWKYVINLGG 229
 Db 180 DRIKESVRLMSQFTVNVVLGKESVLSNGGODPAPHFCCLTILDR--KWDHAILQNF 237
 QY 230 DPELKNFELVSELKLNANMLETVKPPNSKLERFTYHEHLR--VPYEVKLPITRNI 287
 Db 238 DLITKTPYQDSISESLNTSIM-----GFDHGFSYVNNKAKMTPGMKLPIKEIGV 290
 QY 288 SKAPAPRNIIQFVGSAYFLVLSQAFVKYIP---NNSLVQDFPMSKDTYSPDEHFATL 343
 Db 291 PNEILNRNLIVKRSLEVYSKVEFKSMEEKLMDIILIFD-DNDYVGVDEMVLQVLYE 349
 QY 344 RVGIGETSR-BSAQVSDLOQSTRLVKNYIEGFYPSCTGSHLR-SVCIYGAELEW 400
 Db 350 NYLGLGQMSNCTNRNNDI--LTRITMDFSGPNFDECHSKMRHICILMGVEYINME 407
 QY 401 LKDGHEWFKANKDSKVDPIILIKLAEL-----EQORDWIT 437
 Db 408 LIKSQGVIANKVMATFDFGTIACMRMIRMTAGETPNTQWLT 450

RESULT 12
 T32137

Hypothetical protein R07C3.3 - *Caenorhabditis elegans*
 C.Species: *Caenorhabditis elegans*
 C.Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
 C.Accession: J32137
 R:Lamar, B.; Kramer, J.
 Submitted to the EMBL Data Library, July 1997
 A.Description: The sequence of C. elegans cosmid R07C3.
 A.Reference number: 221125
 A.Accession: J32137
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-170 <LAW>
 A.Cross-references: EMBL:AF016686; PIDN:AAB66233.1; GSPDB:GN00020; CESP:R07C3.3
 A.Experimental source: strain Bristol N2; clone R07C3
 C.Genetics:
 A.Gene: CESP:R07C3.3
 A.Map position: 2
 A.Introns: 18/1; 78/3; 189/1; 276/1; 319/3; 359/2; 401/3; 441/3
 C.Superfamily: *Caenorhabditis elegans* hypothetical protein H41C03.3

Query Match	9.2%	Score 219.5;	DB 2;	Length 470;
Best Local Similarity	21.4%	Pred. No. 1.4e-08;		
Matches 95; Conservative	83;	Mismatches 207;	Indels 59;	Gaps 16;

[illegible]

RESULT 13
T21566
Hypothetical protein F30A10.4 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21566
R:Barlow, K.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19442
A:Accession: T21566
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-486 <Wt>
A:Cross-references: EMBL:T21072; PIDN:CA803022.1; GSPDB:GN00019; CESP:F30A10.4
A:Experimental source: Clone F30A10

C:Genetics:
A:Gene: CESP:F30A10.4
A:Map position: 1
A:introns: 19/3; 59/3; 93/2; 248/3; 300/1; 348/3; 385/2; 429/1

Query Match:	9.2%;	Score 219.5;	DB 2;	Length 486;
Best Local Similarity	22.9%;	Pred. No. 1.5e-08;		
Matches 102;	Conservative 80;	Mismatches 202;	Indels 61;	Gaps 17;

```

0Y 15 KEIIFELTLMILSLKLTANV-RLFPQDIY-----LVEYSLSTSEFVRNRYTHK 64
Db 36 REFITIALISTISILLIYYSKMAIPKFKIYRLSGKERAPLKHTRSTGFA--DYFTE 93
0Y 65 DEVRVENCSGIYEOEPLEIKSLDIRRDDIEDDDVYAMTSDCDIYOTLRGYAOKV 144
Db 94 SRSRY-LINCARDIGDVEDISDTYVNNGR--MKIDEKLTQJLSMDCDSIO--NRIFRDM 147
0Y 125 SKEEKSPIAVSLYVHKAIWELIAIYNQHHIYCIHDKRKPDTFKYAMNNLACFS 164
Db 148 PFEKIKRPIAFVRNRYGIYELQEVFLSTISYRPNVFCYAMDSKSEKLSMRIMACFE 207
0Y 185 NIFIASKEAVEVAHISRLQADLCSLDLSSLSQWRVYVILCQDPLKSNPELISELK 244
Db 208 NVITVDEKYMDRGRKHQDAHNPCLKQIDE--HWSHAIITLQNFLLIINSPKLSLSE 265
0Y 245 KINGANML-----ETVYKPNRSKLEERTYTHHELARVYE--YKLPRTNISK 289
Db 266 ILNTYSIGFDYGTFSRYRIFEDWTPAQMKL---FKNE-OSVLEILHKKLIRKSLSE 320
0Y 290 EAPRNIIQFVQSAFVLSQAFVKIETFNNSIVQDFEFAMSKD--TYPSPDEHFATLLI-RVP 346
Db 321 -----VILSKYFVSLPEKVALQNIYKREDDRITLFGVDENMMVMLFENYL 365
0Y 347 GIGETISSAQDVSDLOSRTSLRYLKWNYEG-FYYPSCGSHLR-SVCIYGAELRWLIKD 404
Db 366 GIDGOMESNC-TVAKEDILITQTHHLEBQSGIYQDCKSKMLRHSICVIGIEFLOELSKS 424
0Y 405 GHWFANKPDSKYVPIILIKCLAEKLE 429
Db 425 PVIYANKVTYANPFGTILICVRBMK 449

```

RESULT 14
 T33384
 hypothetical protein HA1C03.3 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T33384
 R:Latreille, P.; Kramer, J. submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of C. elegans cosmid HA1C03.
 A:Reference number: Z21334
 A:Accession: T33384
 A:Status: preliminary; translated from GE/EMBL/DBDJ
 A:Status: preliminary; translated from GE/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-472 <LAI>
 A:Cross-references: EMBL:AF077545; PIDN:AAC26306.1; GSPDB:GND00020; CESP:HA1C03.3
 A:Experimental source: strain Bristol N2; clone HA1C03
 C:Genetics:
 A:Gene: CESP:HA1C03.3
 A:Map position: 2
 A:Intons: 48/1; 94/3; 117/3; 220/1; 307/1; 354/3; 394/2; 406/1; 442/3
 C:Superfamily: *Caenorhabditis elegans* hypothetical protein HA1C03.3

```

Query Match 8.5% Score 203.5 DB 2 Length 472:
Best Local Similarity 20.9% Pred. No. 2, 2e-07:
Matches 90, Conservative 80, Mismatches 189, Indels 71, Gaps 16

QY 20 FTTLLSLSLKLLNVRLF---PQKDIYVERVSLSTSPFVRRNRYHVKDEAYEVN---72
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 19 FKEIWLFLPLAVEVAVYVYIWFENPKKGIY--DPLTIDKQNPPTYGVVDVNDNEIDQRN 75

```

```
QY      20 FTLTWLSLKLINVRRL-----PÖKDIYLVEYSLSGSPFVRRNRTHYKDVERIYN---   72  
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
DB     19 FREIWLLFLAVCFEAYVYLMIFNPKGIY---DPLTIKQNPMWIEQQVVNDLRNELDQRN   75
```

QY 73 -----CSGIYDQEP-----LEIGKSLERIRRDIIIDLEDDDVAMTSDC 110
 Db 76 AISELQKDEYKMKELTEFAIRYPRREPETHIDGRIADKYLTQVSGKDRKIYENK 135
 QY 111 DIYDTLGVAKOKLVSKKEKSPF---TAYSLVHKDAIMVERLIIAHYNOHNTIYCHYDR 166
 Db 136 NLNMSCAIIKMRILPSSNDLKPILKNGVAFARIYTYDEMEIKQVMSYHQNSCFAIDK 195
 QY 167 KAPDTEFVAMNNLAKCSNIFIAASKLEAVEAHISRLQADLNCISDLKSSIQMKYVINTL 226
 Db 196 KAPQFHERLKAACLPNVLPLPDESVDSAGHINILAHNCLRVINKP-GNNVAILL 254
 QY 227 CGDQEPKSNFELVSELKLGANMLETVPKPSKLEPFTYHHELRVPEYKLPF--- 283
 Db 255 QNHDIYKSVYEIQIYDWMLGANDIE-ITPEAGRVDN-----KKFKMDPVSLKXFRN 306
 QY 284 RTINISKEAPPHNIIQIFGSAIFVLSQAIFVKYIFNNSIVQDFP-AMSKDTISPDHEFWATL 342
 Db 307 ETGIDKNVLTTSKMFAGAVQGSLSRGAVDMVMTVDLSTYINQNMDESSGVDEQFIQS- 365
 QY 343 IRVP---GIPGEIS-----RSAQVSDLSQSKTRLVKMY---YEGGFYP--SCTGSHL 387
 Db 366 FQVADLQMGPHFTDECLKOKRNTDFVS-----RMSQMMYGSNKSFDYSIVECTAELL 419
 QY 388 RSVCIYGAEE 397
 Db 420 YNRTFLQVD 429

RESULT 15

T24745
 hypothetical protein T09E11.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 15-Oct-1999
 C:Accession: T24745
 R:McLay, K.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19930
 A:Accession: T24745
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-753 <MIL>
 A:Cross-references: EMBL:Z81147; PDB:CA03539.1; GSPDB:GN00019; CESP:T09E11.6
 A:Experimental source: clone T09E11
 C:Genetics:
 A:Gene: CESP:T09E11.6
 A:Map position: 1
 A:Introns: 34/2; 54/2; 191/3; 275/3; 315/3; 359/3; 380/3; 403/3; 426/3; 530/1; 617/1; 66

Query Match 8.5%; Score 202.5; DB 2; Length 753;

Best Local Similarity 21.8%; Pred. No. 4.9e-07;
 Matches 96; Conservative 76; Mismatches 207; Indels 61; Gaps 16;

QY 21 LTLWL-----SLKLINVRRLFPQKDIY--LVEYSLS-----TSPFVNRRTYHWKDEV 67
 Db 335 LNRWLILIIAASFISYFLVILLEKEKEPYQTLSSTLSNCEGPTKODLRHHTDIADY 394
 QY 68 R---YEVNCSGIYDQEP-----LEIGKSLERIRRDIIIDLEDDDVAMTSDCDIYQTLRGYA 120
 Db 395 RERGFADDKRQSYRRREPETHAVDCGRILAGDKPYLQSLTGTNRKTYENCNLNMSCFAIR 454
 QY 121 QKLVSKEEK-----SPFIAYSLVHKDAIMVERLIIAHYNOHNTIYCHYDRKAPDTEKVA 175
 Db 455 SRLPSNDNILRPLKGIARFYKDYEFIERQVQVSFPQNAFCFVIDINMSEFKR 514
 QY 176 MNNLAKCFNSNFIASKLEAVEYAHISRLQADLNCISDLKSSIQMKYVINTLCCGDPPLKS 235
 Db 515 MRALAACMPNIVYIADDEDPYSSGHNVLVHNKCKALLDIP-GNNYALLQNHNDLTKMS 573
 QY 236 NFELVSELKLGANMLETVPKPSKLEPFTYHHELRVPEYKLPF-----RTNI 287

Db 574 VYEMQIFEMLGANDIEVT-----HEIGRYDVKKLKMPPSKLFINETEM 620
 QY 288 SKAAPPHNIIQIFVGSAYFVLSQAFVKYIFNNSIVQDFP-AMSKDTISPDHEFWATLIRVP 346
 Db 621 DKLLTTPMKIYKGMVHCSLSRASVEMFQKLDPSIFMHQLNQGRYGVDEQYFPILOANA 680
 QY 347 --GIPGEIS-RSAQVSDLSQSKTRLVKMYYEGGFYP--SC-TGSHLRVCIYGAELRW 400
 Db 681 EFGMPGHFTDECIQSGKTEFTFIRILW-----VPESKDTNMTNRRAVCIIGLEHPQA 733
 QY 401 LKDGHWFRANKPDSKYDPIL 420
 Db 734 VASFTLMEFNKVSSSLDELI 753

Search completed: May 2, 2002, 07:45:39
 Job time: 150 sec

Fri May 3 10:57:56 2002

us-09-645-192-2.rpt

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 11:37:39 ; Search time 172.38 Seconds

(without alignments)
6773.854 Million cell updates/sec

Title: US-09-645-192-1

Perfect score: 1362
Sequence: 1 atgaagatcatcaatgta.....atctcactcaccacatcatga 1362

Scoring table: OLIGO-NUC
Gapop 60.0, Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size: 10

Total number of hits satisfying chosen parameters: 456299

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database: N_Geneseq.1101.*

```
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1362	100.0	1362	22	AA500045	Human DNA encoding
2	349	25.6	781	22	AA98678	Human EST-derived
3	183	13.4	186	21	AA19209	Human secreted pro
4	25	1.8	31	22	AA500048	Human C2Gnt3 PCR p
5	5	1.5	21	21	AA500046	Human C2Gnt3 PCR p
6	20	1.5	32	22	AA500047	Human C2Gnt3 PCR p
7	20	1.5	1086	15	AA052335	Human C2Gnt3 PCR p
8	19	1.4	364	21	AA331543	Influenza type B H
9	19	1.4	1127	21	AA41094	Plant microsatelli
10	18	1.3	20	22	AA500049	Arabidopsis thalia
11	18	1.3	255	22	AAH81918	Human C2Gnt3 PCR p
						Ret differential t

12	13	300	20	AA213168	Human gene express
13	13	528	21	AA118103	Aspergillus niger
14	13	592	20	AA12009	Neisseria meningit
15	13	592	21	AA81331	N. meningitidis Me
16	13	807	20	AA12010	Complete ORF17 seq
17	13	807	20	AA12011	Complete ORF17 seq
18	13	807	20	AA12012	Complete ORF17 seq
19	13	1160	18	AA193294	Tomato mottle viru
20	13	1166	18	AA193285	Tomato mottle viru
21	13	1169	18	AA193282	Tomato mottle viru
22	13	1169	18	AA193283	Tomato mottle viru
23	13	1463	18	AA193284	Tomato mottle viru
24	13	2456	21	AA422240	Arabidopsis thalia
25	13	2602	18	AA193309	N. meningitidis pa
26	13	2843	21	AA193309	Tomato mottle viru
27	13	2843	21	AA193309	Protein regulating
28	13	3864	22	AA144664	Novel protein kina
29	13	4982	21	AA175062	Human ORF17 ORF617
30	13	5355	20	AA189220	Seq ID No: 28 of W
31	13	5355	20	AA189221	Seq ID No: 30 of W
32	13	5355	21	AA189221	Human TAD protein
33	13	5355	21	AA189221	Human glycosyl sul
34	13	5355	21	AA189221	Human T gene DNA.
35	13	5355	21	AA189221	Human T gene DNA.
36	13	5355	21	AA189221	Human T gene DNA.
37	13	5355	21	AA189221	Human T gene DNA.
38	13	5355	21	AA189221	Human T gene DNA.
39	13	5355	21	AA189221	Human T gene DNA.
40	13	5355	21	AA189221	Human T gene DNA.
41	13	5355	21	AA189221	Human T gene DNA.
42	13	5355	21	AA189221	Human T gene DNA.
43	13	5355	21	AA189221	Human T gene DNA.
44	13	5355	21	AA189221	Human T gene DNA.
45	13	5355	21	AA189221	Human T gene DNA.
46	13	5355	21	AA189221	Human T gene DNA.
47	13	5355	21	AA189221	Human T gene DNA.
48	13	5355	21	AA189221	Human T gene DNA.
49	13	5355	21	AA189221	Human T gene DNA.
50	13	5355	21	AA189221	Human T gene DNA.
51	13	5355	21	AA189221	Human T gene DNA.
52	13	5355	21	AA189221	Human T gene DNA.
53	13	5355	21	AA189221	Human T gene DNA.
54	13	5355	21	AA189221	Human T gene DNA.
55	13	5355	21	AA189221	Human T gene DNA.
56	13	5355	21	AA189221	Human T gene DNA.
57	13	5355	21	AA189221	Human T gene DNA.
58	13	5355	21	AA189221	Human T gene DNA.
59	13	5355	21	AA189221	Human T gene DNA.
60	13	5355	21	AA189221	Human T gene DNA.
61	13	5355	21	AA189221	Human T gene DNA.
62	13	5355	21	AA189221	Human T gene DNA.
63	13	5355	21	AA189221	Human T gene DNA.
64	13	5355	21	AA189221	Human T gene DNA.
65	13	5355	21	AA189221	Human T gene DNA.
66	13	5355	21	AA189221	Human T gene DNA.
67	13	5355	21	AA189221	Human T gene DNA.
68	13	5355	21	AA189221	Human T gene DNA.
69	13	5355	21	AA189221	Human T gene DNA.
70	13	5355	21	AA189221	Human T gene DNA.
71	13	5355	21	AA189221	Human T gene DNA.
72	13	5355	21	AA189221	Human T gene DNA.
73	13	5355	21	AA189221	Human T gene DNA.
74	13	5355	21	AA189221	Human T gene DNA.
75	13	5355	21	AA189221	Human T gene DNA.
76	13	5355	21	AA189221	Human T gene DNA.
77	13	5355	21	AA189221	Human T gene DNA.
78	13	5355	21	AA189221	Human T gene DNA.
79	13	5355	21	AA189221	Human T gene DNA.
80	13	5355	21	AA189221	Human T gene DNA.
81	13	5355	21	AA189221	Human T gene DNA.
82	13	5355	21	AA189221	Human T gene DNA.
83	13	5355	21	AA189221	Human T gene DNA.
84	13	5355	21	AA189221	Human T gene DNA.

C 85	17	1.2	37940	20	AA201026	Partial mouse PGI
C 86	17	1.2	50000	22	AAF54867	Nucleotide sequenc
C 87	17	1.2	53585	20	AA220251	Borrelia burgdorferi
C 88	17	1.2	80450	21	AAF22295	BAC containing rep
C 89	17	1.2	81145	22	AAF54868	Genomic nucleotide
C 90	17	1.2	130480	22	AAF25833	R. marinus bacteri
C 91	17	1.2	160552	22	AA020697	Human glycosyl sul
C 92	17	1.2	349980	22	AAH68532	C glutamicum codin
C 93	17	1.2	349980	22	AAH68533	C glutamicum codin
C 94	17	1.2	1664976	19	AAV21209	Methanococcus jann
C 95	16	1.2		20	AA083732	primer D8, to gene
C 96	16	1.2		21	AAH40578	SNP specific lower
C 97	16	1.2		22	AAH01917	biatrem resistance
C 98	16	1.2		24	AAV62884	Primer #3 for Full
C 99	16	1.2		24	AAV82610	Oligo-Amp used for
C 100	16	1.2		21	AAV99117	Plasmid p773-Pac
C 101	16	1.2		27	AAAT48159	Escherichia coli c
C 102	16	1.2		28	AAA37855	terramyl-like alph
C 103	16	1.2		30	AAQ1868	scat closing oligo
C 104	16	1.2		30	AAQ1871	Closing oligonucle
C 105	16	1.2		30	AAV32985	PCR primer NLC430
C 106	16	1.2		30	AAV23741	Cellulase variant
C 107	16	1.2		30	AAV27950	PCR primer for H.
C 108	16	1.2		31	AA500052	Human CG2nt3 seque
C 109	16	1.2		34	AA60677	MLV variant rever
C 110	16	1.2		40	AA68440	Plasmid p18Sflr c
C 111	16	1.2		40	AA68440	Circular plasmid e
C 112	16	1.2		55	AAK01785	Human cystic fibro
C 113	16	1.2		100	AAAT3089	Primer 26 for 95 K
C 114	16	1.2		111	AAAC14092	Human secreted pro
C 115	16	1.2		114	AAAC14952	Human secreted pro
C 116	16	1.2		121	AAI21554	Probe #11487 for g
C 117	16	1.2		121	AAI46845	Probe #15531 used t
C 118	16	1.2		121	AAI07248	Probe #7239 used t
C 119	16	1.2		167	AAV88342	EST clone GA327.
C 120	16	1.2		169	AAA41761	Human secreted exp
C 121	16	1.2		180	AAV89351	EST clone C1415.
C 122	16	1.2		190	AAI21532	Probe #11465 for g
C 123	16	1.2		190	AAI46824	Probe #15510 used
C 124	16	1.2		229	AAI07228	Probe #15510 used
C 125	16	1.2		229	AAH50280	Probe #7219 used t
C 126	16	1.2		229	AAH50335	Bacterial 23S/5S R
C 127	16	1.2		237	AAAC28394	Human secreted pro
C 128	16	1.2		279	AAI21667	Probe #11600 for g
C 129	16	1.2		279	AAI46956	Probe #15642 used
C 130	16	1.2		279	AAI07358	Probe #7349 used t
C 131	16	1.2		287	AAI23003	Probe #15980 used
C 132	16	1.2		287	AAI47294	Probe #15980 used
C 133	16	1.2		287	AAI07699	Probe #7690 used t
C 134	16	1.2		288	AAI20039	Human gene signatu
C 135	16	1.2		291	AAI52671	Probe #21357 used
C 136	16	1.2		300	AAAC23456	Human secreted pro
C 137	16	1.2		300	AAA00029	Human colon cancer
C 138	16	1.2		302	AAV40308	Human secreted pro
C 139	16	1.2		307	AAA43540	Human secreted exp
C 140	16	1.2		316	AAA44328	Human secreted exp
C 141	16	1.2		324	AAH93972	Human foetal cDNA,
C 142	16	1.2		339	AAV41125	Human secreted pro
C 143	16	1.2		342	AAI63085	A diacylglycerol a
C 144	16	1.2		350	AAI63064	Human kidney relat
C 145	16	1.2		353	AAV88346	EST clone GA205.
C 146	16	1.2		381	AAAT67613	H. pylori cytoplas
C 147	16	1.2		386	AAV64286	Novel human polyu
C 148	16	1.2		387	AAV21955	Human breast and c
C 149	16	1.2		410	AAV20320	Probe (221) for m
C 150	16	1.2		427	AAV79304	Human lung tumour-
C 151	16	1.2		437	AAH72101	Human cervical can
C 152	16	1.2		451	AAI12475	Probe #2408 for ge
C 153	16	1.2		451	AAI33830	Probe #2516 used t
C 154	16	1.2		451	AAI03386	Probe #2377 used t
C 155	16	1.2		455	AAI04414	Probe #4405 used t
C 156	16	1.2		457	AAI21172	Polynucleotide seq
C 157	16	1.2		461	AAI11944	Probe #1877 for ge
C 158	16	1.2		461	AAI33276	Probe #1962 used t
C 159	16	1.2		461	AAI01877	Probe #642 for gen
C 160	16	1.2		468	AAI10709	Probe #4028 for ge
C 161	16	1.2		468	AAI14095	Probe #653 used to
C 162	16	1.2		468	AAI31967	Probe #4162 used t
C 163	16	1.2		468	AAI35476	Probe #630 used to
C 164	16	1.2		468	AAI00639	Probe #3938 used t
C 165	16	1.2		469	AAI03947	Human secreted pro
C 166	16	1.2		469	AAI00988	Probe #17070 for g
C 167	16	1.2		470	AAI27337	Probe #24674 used
C 168	16	1.2		470	AAI55968	Probe #118 for gen
C 169	16	1.2		478	AAI10185	Probe #1120 used to
C 170	16	1.2		478	AAI31434	Probe #118 used to
C 171	16	1.2		478	AAI00127	CDNA encoding SRT
C 172	16	1.2		478	AAV93487	Human cervical can
C 173	16	1.2		479	AAV69553	Probe #2270 for ge
C 174	16	1.2		480	AAI12337	Probe #2379 used t
C 175	16	1.2		480	AAI33653	Probe #2242 used t
C 176	16	1.2		480	AAI02251	Human cervical can
C 177	16	1.2		489	AAH72422	Human cervical can
C 178	16	1.2		489	AAH72990	Human cervical can
C 179	16	1.2		492	AAH70868	S. aureus phosphor
C 180	16	1.2		522	AAV26098	Human secreted pro
C 181	16	1.2		524	AAAC01027	S. aureus phosphor
C 182	16	1.2		525	AAV26097	Polynucleotide seq
C 183	16	1.2		525	AAV21136	DNA encoding a Sta
C 184	16	1.2		537	AAH83938	Drosophila melanog
C 185	16	1.2		558	AAH29611	Probe #9605 used t
C 186	16	1.2		559	AAI40919	Human normal pancr
C 187	16	1.2		560	AAZ41409	Probe #7882 for ge
C 188	16	1.2		587	AAI17949	Probe #11623 used
C 189	16	1.2		587	AAI42937	Human CDNA clone (
C 190	16	1.2		589	AAH09408	Probe #11219 used
C 191	16	1.2		594	AAI42533	Probe #8275 used t
C 192	16	1.2		598	AAI39589	Probe #7515 for ge
C 193	16	1.2		600	AAI17582	Probe #11183 used
C 194	16	1.2		600	AAI42497	PRSV field strain
C 195	16	1.2		603	AAV79565	PRSV field strain
C 196	16	1.2		603	AAV79560	PRSV field strain
C 197	16	1.2		603	AAV79557	PRSV field strain
C 198	16	1.2		603	AAV79568	PRSV field strain
C 199	16	1.2		605	AAV53473	DNA encoding a Sta
C 200	16	1.2		618	AAA43080	Human secreted exp
C 201	16	1.2		648	AAH77317	IgG Fab-BPI fusion
C 202	16	1.2		654	AAV60631	Beta-urogastrone -
C 203	16	1.2		658	AAV60631	Human BPI CDNA seq
C 204	16	1.2		661	AAV08242	Fusarium venenatum
C 205	16	1.2		664	AAH71030	Human cervical can
C 206	16	1.2		667	AAH06953	Human CDNA clone (
C 207	16	1.2		699	AAZ80206	Human colon cancer
C 208	16	1.2		702	AAH07852	Human CDNA clone (
C 209	16	1.2		720	AAV13962	Human immunoblan
C 210	16	1.2		728	AAV30915	Human EST-derived
C 211	16	1.2		734	AAH98602	Yeast 2,5-Diamino-
C 212	16	1.2		735	AAH40401	Human gene express
C 213	16	1.2		765	AAI16056	Human secreted pro
C 214	16	1.2		768	AAV05339	Human colon cancer
C 215	16	1.2		769	AAA01709	Human prostate can
C 216	16	1.2		770	AAZ97293	Human prostate can
C 217	16	1.2		771	AAZ97293	Human prostate can
C 218	16	1.2		776	AAV22989	Human CDNA clone (
C 219	16	1.2		776	AAH06486	Human CDNA clone (
C 220	16	1.2		779	AAH08643	Human CDNA clone (
C 221	16	1.2		779	AAV36964	Human pituitary-tu
C 222	16	1.2		780	AAZ97292	Human prostate can
C 223	16	1.2		789	AAI10411	E. coli mature TEM
C 224	16	1.2		792	AAV090773	Escherichia coli m
C 225	16	1.2		792	AAV37842	Escherichia coli m
C 226	16	1.2		792	AAV17986	Escherichia coli m
C 227	16	1.2		792	AAV32377	RTM with a mammal
C 228	16	1.2		792	AAI33046	E. coli TEM-1 beta
C 229	16	1.2		795	AAV37859	E. coli RTM beta-
C 230	16	1.2		795	AAV17983	Escherichia coli m
					AAV32374	Escherichia coli R

231	1.2	795	22	AAD13043	E. coli TEM-1 beta	304	1.2	1366	22	AAH15377	Human cDNA sequenc
232	1.6	795	22	AAD13045	E. coli TEM-1 beta	305	1.6	1366	11	AAO62236	Amb a II clone enc
233	1.2	798	22	AAH03926	Human cDNA clone (306	1.2	1368	19	AAV37864	Allergen Amb a II
234	1.2	803	22	AAH06590	Human cDNA clone (307	1.6	1368	19	AAV13361	Ragweed pollen Amb
235	1.2	808	22	AAH01936	Klebsiella oxyloca	308	1.2	1385	21	AAc42253	Arabidopsis thalia
236	1.2	817	19	AAV40858	Beta-lactamase fra	309	1.6	1392	22	AAO50340	Human secreted pro
237	1.2	817	22	AAH07959	Human cDNA clone (310	1.2	1395	14	AAO51243	Ragweed Pollen All
238	1.2	819	19	AAV63198	cDNA from clone ga	311	1.2	1395	14	AAV32618	Ragweed Amb all al
239	1.6	826	21	AAc44565	2ea mays DNA fragm	312	1.6	1403	21	AAc36040	Arabidopsis thalia
240	1.2	827	21	AAc61643	Human hair specific	313	1.2	1425	20	AAx20727	Polynucleotide seq
241	1.2	835	20	AAx57719	DNA encoding a hum	314	1.2	1426	20	AAx20336	Botrellia burgotte
242	1.6	840	19	AAV17985	Escherichia coli m	315	1.6	1433	21	AAc41836	Arabidopsis thalia
243	1.2	843	17	AAx37841	E. coli RTM beta-	316	1.2	1443	21	AAc40612	Arabidopsis thalia
244	1.2	843	17	AAx37841	RTM with beta-glo	317	1.2	1464	21	AAc41677	Arabidopsis thalia
245	1.2	843	22	AAH06590	Human cDNA clone (318	1.2	1465	22	AAI60352	Human polynucleoti
246	1.2	858	17	AAx37840	E. coli RTM beta-	319	1.2	1469	22	AAH42032	Disease treatment
247	1.2	858	19	AAV17984	Escherichia coli m	320	1.6	1470	22	AAI58566	Human polynucleoti
248	1.2	858	20	AAz33375	Escherichia coli R	321	1.2	1473	20	AAI28825	Human polynucleoti
249	1.2	858	22	AAI37044	E. coli TEM-1 beta	322	1.2	1488	21	AAc37961	Rat neuronal lamed
250	1.2	860	22	AAH06593	Human cDNA clone (323	1.2	1496	20	AAI28233	Arabidopsis thalia
251	1.2	861	17	AAI28560	Bacterial antidiot	324	1.2	1542	17	AAI42305	Human calcium bind
252	1.6	861	17	AAI31627	Tem gene encodes b	325	1.2	1542	17	AAI42305	Male-specific DNA
253	1.2	861	22	AAH01934	Neisseria meningit	326	1.2	1546	22	AAc84564	Nucleotide sequenc
254	1.2	861	22	AAH01937	Escherichia coli n	327	1.2	1598	22	AAO55306	Human secreted pro
255	1.2	861	22	AAH01938	Escherichia coli n	328	1.2	1602	21	AAO7664	Fusarium venenatum
256	1.2	861	22	AAH01939	Escherichia coli n	329	1.2	1610	20	AAx13498	Enterococcus faeca
257	1.2	861	22	AAH01940	Escherichia coli n	330	1.2	1616	21	AAc50714	Arabidopsis thalia
258	1.2	861	22	AAH02018	Klebsiella coli n	331	1.2	1618	21	AAc39869	Arabidopsis thalia
259	1.2	864	22	AAH00262	Bartonella hensela	332	1.2	1653	19	AAV39861	Bactericidal perme
260	1.2	869	22	AAH72931	Human cervical can	333	1.2	1662	22	AAH02237	Tetragonococcus ha
261	1.2	870	17	AAI07318	Partial formyl-CoA	334	1.2	1666	22	AAH15408	Human cDNA sequenc
262	1.2	870	17	AAV64282	O. formigenes form	335	1.2	1692	21	AAI25375	Human secreted pro
263	1.2	877	21	AAx44356	Human secreted exp	336	1.2	1707	21	AAc39785	Arabidopsis thalia
264	1.2	900	19	AAV3860	Bactericidal perme	337	1.2	1722	21	AAH16832	Human cDNA sequenc
265	1.2	908	20	AAx89295	Human regulatory p	338	1.2	1727	21	AAz47340	DDC3 (dopa decarbo
266	1.2	933	22	AAH31708	Human olfactory re	339	1.2	1729	21	AAc36534	Arabidopsis thalia
267	1.2	933	22	AAH31768	Human olfactory re	340	1.2	1768	20	AAx37405	Human secreted pro
268	1.2	948	21	AAc48497	Arabidopsis thalia	341	1.2	1799	18	AAI59214	Influenza B/Panama
269	1.2	965	19	AAV43035	Streptococcus pneu	342	1.2	1799	22	AAx00775	Influenza virus B/
270	1.2	975	21	AAc42939	Arabidopsis thalia	343	1.2	1799	22	AAO09568	Influenza virus B/
271	1.2	1000	22	AAI94412	N. meningitidis (s	344	1.2	1802	18	AAI59218	Influenza virus B/
272	1.2	1037	21	AAc58721	Human secreted pro	345	1.2	1802	22	AAx00779	Influenza B/Hardin
273	1.2	1075	21	AAc34728	Arabidopsis thalia	346	1.2	1802	22	AAO09592	Influenza virus B/
274	1.2	1080	15	AAO62337	Influenza type B H	347	1.2	1810	13	AAQ25593	Sequence encoding
275	1.2	1080	15	AAO62338	Influenza type B H	348	1.2	1811	18	AAI59215	Influenza B/Nether
276	1.2	1080	15	AAO62339	Influenza type B H	349	1.2	1811	20	AAx00776	Influenza virus B/
277	1.2	1080	15	AAO62340	Influenza type B H	350	1.2	1811	22	AAO09589	Influenza virus B/
278	1.2	1080	15	AAO62341	Influenza type B H	351	1.2	1813	10	AAI92766	CDNA encoding huma
279	1.2	1082	21	AAc39923	Arabidopsis thalia	352	1.2	1813	14	AAQ52268	Human bactericidal
280	1.2	1086	15	AAO62334	Influenza type B H	353	1.2	1813	14	AAQ52268	Human bactericidal
281	1.2	1086	15	AAO62347	Influenza type B H	354	1.2	1813	15	AAQ67270	pinG4512 coding re
282	1.2	1106	14	AAO49225	Plasmid pDS56/RBSI	355	1.2	1813	15	AAQ73382	Recombinant bacter
283	1.2	1106	14	AAO49226	Plasmid pDS56/RBSI	356	1.2	1813	15	AAO72017	CDNA encoding reco
284	1.2	1106	14	AAO49226	Plasmid pDS56/RBSI	357	1.2	1813	15	AAO09551	Bactericidal/perme
285	1.2	1113	21	AAI5981	Human prostate can	358	1.2	1813	16	AAO86285	Gene encoding rpPI
286	1.2	1125	21	AAc48273	Arabidopsis thalia	359	1.2	1813	16	AAO99816	Recombinant bacter
287	1.2	1130	21	AAc40921	Arabidopsis thalia	360	1.2	1813	16	AAO85883	Human bactericidal
288	1.2	1144	22	AAH29580	Drosophila melanog	361	1.2	1813	16	AAO81445	Holoolein rBP150
289	1.2	1163	18	AAI87973	PCAL module M10-ii	362	1.2	1813	16	AAO80827	Bactericidal/perme
290	1.2	1167	21	AAAS8885	Arabidopsis Est sh	363	1.2	1813	16	AAO92640	Recombinant bacter
291	1.2	1169	22	AAI59058	Human polynucleoti	364	1.2	1813	16	AAO92640	DNA encoding recom
292	1.2	1212	5	AAH40167	Sequence encoding	365	1.2	1813	16	AAO97607	Bactericidal/perme
293	1.2	1237	21	AAc40215	Arabidopsis thalia	366	1.2	1813	17	AAI31853	Bactericidal/perme
294	1.2	1238	21	AAc45544	Arabidopsis thalia	367	1.2	1813	17	AAI31853	Bactericidal/perme
295	1.2	1256	22	AAH29938	C. albicans apoptos	368	1.2	1813	18	AAV06404	Recombinant bacter
296	1.2	1258	18	AAI91739	DNA gyrase subunit	369	1.2	1813	18	AAI67139	Human bactericidal
297	1.2	1259	21	AAc42410	Arabidopsis thalia	370	1.2	1813	18	AAI67140	Bactericidal/perme
298	1.2	1266	14	AAH53212	S. epidermidis ope	371	1.2	1813	19	AAV38504	CDNA encoding huma
299	1.2	1273	14	AAO42410	Tripartite fusion	372	1.2	1813	19	AAV36465	Recombinant BPI CD
300	1.2	1289	18	AAI88300	Beta-lactamase-WCS	373	1.2	1813	19	AAI22675	CDNA encoding a ba
301	1.2	1291	21	AAc40588	Arabidopsis thalia	374	1.2	1813	19	AAI99564	Human bactericidal
302	1.2	1305	21	AAc51959	Arabidopsis thalia	375	1.2	1813	19	AAV10337	Human recombinant
303	1.2	1334	22	AAH15385	Human cDNA sequenc	376	1.2	1813	19	AAV14996	Human bactericidal
										AAV13961	Human BPI protein

C 377	16	1.2	1813	20	AA224943	Recombinant human	C 450	16	1.2	2557	18	AA768842	Photornhabus lumini
C 378	16	1.2	1813	20	AA211725	Human bactericidal	C 451	16	1.2	2557	19	AAV29929	Internal gene frag
C 379	16	1.2	1813	20	AA280157	Human bactericidal	C 452	16	1.2	2573	16	AAQ79138	Nas gene derived f
C 380	16	1.2	1813	20	AA118600	CDNA encoding huma	C 453	16	1.2	2577	14	AAQ36620	Expression plasmid
C 381	16	1.2	1813	20	AAAX05716	Human bactericidal	C 454	16	1.2	2583	14	AAQ46606	Plasmid pINT41d co
C 382	16	1.2	1813	21	AAAS44281	Bactericidal/perme	C 455	16	1.2	2681	17	AA130870	Engineered 95 kD p
C 383	16	1.2	1813	21	AA4756800	DNA encoding a hum	C 456	16	1.2	2686	18	AA161831	Puc18. Synthetic.
C 384	16	1.2	1813	21	AAAS8978	DNA encoding a hum	C 457	16	1.2	2685	22	AA459053	Plasmid vector pUC
C 385	16	1.2	1813	21	AA462318	Human recombinant	C 458	16	1.2	2698	22	AA428230	Nucleotide sequenc
C 386	16	1.2	1813	21	AA462832	Recombinant holopr	C 459	16	1.2	2700	20	AA429680	C. elegans kinase
C 387	16	1.2	1813	21	AA413265	Human bactericidal	C 460	16	1.2	2704	20	AA206758	Vector pUC28 nucle
C 388	16	1.2	1813	21	AA257575	Human bactericidal	C 461	16	1.2	2728	18	AA187983	Vector pUC28 modu
C 389	16	1.2	1813	21	AA2444274	Human bactericidal	C 462	16	1.2	2731	21	AA414802	Vector pGEMT for e
C 390	16	1.2	1813	22	AA500568	Human BPI DNA. Ho	C 463	16	1.2	2739	20	AA401067	Vector pGEMT DNA se
C 391	16	1.2	1813	22	AA501832	Human recombinant	C 464	16	1.2	2747	22	AA401067	Human CDNA sequenc
C 392	16	1.2	1813	22	AA508052	Human BPI variant	C 465	16	1.2	2753	22	AA475001	Human CDNA sequenc
C 393	16	1.2	1813	22	AA508541	Human BPI protein	C 466	16	1.2	2754	21	AA475008	Human CDNA sequenc
C 394	16	1.2	1813	22	AA508542	Human BPI protein	C 467	16	1.2	2755	21	AA475008	Human CDNA sequenc
C 395	16	1.2	1813	22	AA508542	DNA encoding a hum	C 468	16	1.2	2755	13	AA403056	Vector pPUC3. Syn
C 396	16	1.2	1813	22	AA508542	Human BPI coding s	C 469	16	1.2	2838	22	AA403148	Vector pPUC3. Syn
C 397	16	1.2	1813	22	AA508542	Human BPI coding s	C 470	16	1.2	2856	22	AA403148	Expression vector
C 398	16	1.2	1813	22	AA508542	Human BPI coding s	C 471	16	1.2	2869	22	AA403148	Human CDNA sequenc
C 399	16	1.2	1813	22	AA508542	Human BPI coding s	C 472	16	1.2	2870	22	AA403148	Expression vector
C 400	16	1.2	1813	22	AA508542	Human BPI coding s	C 473	16	1.2	2871	19	AA403148	Nucleotide sequenc
C 401	16	1.2	1813	22	AA508542	Human BPI coding s	C 474	16	1.2	2871	19	AA403148	Expression vector
C 402	16	1.2	1813	22	AA508542	Human BPI coding s	C 475	16	1.2	2871	19	AA403148	Expression vector
C 403	16	1.2	1813	22	AA508542	Human BPI coding s	C 476	16	1.2	2871	19	AA403148	Expression vector
C 404	16	1.2	1813	22	AA508542	Human BPI coding s	C 477	16	1.2	2871	19	AA403148	Expression vector
C 405	16	1.2	1813	22	AA508542	Human BPI coding s	C 478	16	1.2	2871	19	AA403148	Expression vector
C 406	16	1.2	1813	22	AA508542	Human BPI coding s	C 479	16	1.2	2871	19	AA403148	Expression vector
C 407	16	1.2	1813	22	AA508542	Human BPI coding s	C 480	16	1.2	2871	19	AA403148	Expression vector
C 408	16	1.2	1813	22	AA508542	Human BPI coding s	C 481	16	1.2	2871	19	AA403148	Expression vector
C 409	16	1.2	1813	22	AA508542	Human BPI coding s	C 482	16	1.2	2871	19	AA403148	Expression vector
C 410	16	1.2	1813	22	AA508542	Human BPI coding s	C 483	16	1.2	2871	19	AA403148	Expression vector
C 411	16	1.2	1813	22	AA508542	Human BPI coding s	C 484	16	1.2	2871	19	AA403148	Expression vector
C 412	16	1.2	1813	22	AA508542	Human BPI coding s	C 485	16	1.2	2871	19	AA403148	Expression vector
C 413	16	1.2	1813	22	AA508542	Human BPI coding s	C 486	16	1.2	2871	19	AA403148	Expression vector
C 414	16	1.2	1813	22	AA508542	Human BPI coding s	C 487	16	1.2	2871	19	AA403148	Expression vector
C 415	16	1.2	1813	22	AA508542	Human BPI coding s	C 488	16	1.2	2871	19	AA403148	Expression vector
C 416	16	1.2	1813	22	AA508542	Human BPI coding s	C 489	16	1.2	2871	19	AA403148	Expression vector
C 417	16	1.2	1813	22	AA508542	Human BPI coding s	C 490	16	1.2	2871	19	AA403148	Expression vector
C 418	16	1.2	1813	22	AA508542	Human BPI coding s	C 491	16	1.2	2871	19	AA403148	Expression vector
C 419	16	1.2	1813	22	AA508542	Human BPI coding s	C 492	16	1.2	2871	19	AA403148	Expression vector
C 420	16	1.2	1813	22	AA508542	Human BPI coding s	C 493	16	1.2	2871	19	AA403148	Expression vector
C 421	16	1.2	1813	22	AA508542	Human BPI coding s	C 494	16	1.2	2871	19	AA403148	Expression vector
C 422	16	1.2	1813	22	AA508542	Human BPI coding s	C 495	16	1.2	2871	19	AA403148	Expression vector
C 423	16	1.2	1813	22	AA508542	Human BPI coding s	C 496	16	1.2	2871	19	AA403148	Expression vector
C 424	16	1.2	1813	22	AA508542	Human BPI coding s	C 497	16	1.2	2871	19	AA403148	Expression vector
C 425	16	1.2	1813	22	AA508542	Human BPI coding s	C 498	16	1.2	2871	19	AA403148	Expression vector
C 426	16	1.2	1813	22	AA508542	Human BPI coding s	C 499	16	1.2	2871	19	AA403148	Expression vector
C 427	16	1.2	1813	22	AA508542	Human BPI coding s	C 500	16	1.2	2871	19	AA403148	Expression vector
C 428	16	1.2	1813	22	AA508542	Human BPI coding s	C 501	16	1.2	2871	19	AA403148	Expression vector
C 429	16	1.2	1813	22	AA508542	Human BPI coding s	C 502	16	1.2	2871	19	AA403148	Expression vector
C 430	16	1.2	1813	22	AA508542	Human BPI coding s	C 503	16	1.2	2871	19	AA403148	Expression vector
C 431	16	1.2	1813	22	AA508542	Human BPI coding s	C 504	16	1.2	2871	19	AA403148	Expression vector
C 432	16	1.2	1813	22	AA508542	Human BPI coding s	C 505	16	1.2	2871	19	AA403148	Expression vector
C 433	16	1.2	1813	22	AA508542	Human BPI coding s	C 506	16	1.2	2871	19	AA403148	Expression vector
C 434	16	1.2	1813	22	AA508542	Human BPI coding s	C 507	16	1.2	2871	19	AA403148	Expression vector
C 435	16	1.2	1813	22	AA508542	Human BPI coding s	C 508	16	1.2	2871	19	AA403148	Expression vector
C 436	16	1.2	1813	22	AA508542	Human BPI coding s	C 509	16	1.2	2871	19	AA403148	Expression vector
C 437	16	1.2	1813	22	AA508542	Human BPI coding s	C 510	16	1.2	2871	19	AA403148	Expression vector
C 438	16	1.2	1813	22	AA508542	Human BPI coding s	C 511	16	1.2	2871	19	AA403148	Expression vector
C 439	16	1.2	1813	22	AA508542	Human BPI coding s	C 512	16	1.2	2871	19	AA403148	Expression vector
C 440	16	1.2	1813	22	AA508542	Human BPI coding s	C 513	16	1.2	2871	19	AA403148	Expression vector
C 441	16	1.2	1813	22	AA508542	Human BPI coding s	C 514	16	1.2	2871	19	AA403148	Expression vector
C 442	16	1.2	1813	22	AA508542	Human BPI coding s	C 515	16	1.2	2871	19	AA403148	Expression vector
C 443	16	1.2	1813	22	AA508542	Human BPI coding s	C 516	16	1.2	2871	19	AA403148	Expression vector
C 444	16	1.2	1813	22	AA508542	Human BPI coding s	C 517	16	1.2	2871	19	AA403148	Expression vector
C 445	16	1.2	1813	22	AA508542	Human BPI coding s	C 518	16	1.2	2871	19	AA403148	Expression vector
C 446	16	1.2	1813	22	AA508542	Human BPI coding s	C 519	16	1.2	2871	19	AA403148	Expression vector
C 447	16	1.2	1813	22	AA508542	Human BPI coding s	C 520	16	1.2	2871	19	AA403148	Expression vector
C 448	16	1.2	1813	22	AA508542	Human BPI coding s	C 521	16	1.2	2871	19	AA403148	Expression vector
C 449	16	1.2	1813	22	AA508542	Human BPI coding s	C 522	16	1.2	2871	19	AA403148	Expression vector

C 523	16	1.2	3332	21	AAQ66453	Human secreted pro	C 596	16	1.2	3737	19	AAV43796	Sequence ID No.1 f
C 524	16	1.2	3343	11	AAQ04655	Plasmid p10163 enc	C 597	16	1.2	3752	22	AAV55130	Nucleotide sequenc
C 525	16	1.2	3344	11	AAQ39495	Transgenic unc-119	C 598	16	1.2	3754	16	AAQ57880	Intermediate plasm
C 526	16	1.2	3363	20	AAZ08715	Alpha hordochthonin	C 599	16	1.2	3754	16	AAQ85420	Plasmid pRECH. S
C 527	16	1.2	3365	20	AAZ08716	Alpha hordochthonin	C 600	16	1.2	3769	15	AAQ57881	Intermediate plasm
C 528	16	1.2	3369	22	AAQ86255	PGHRH-4 44SK const	C 601	16	1.2	3769	16	AAQ97492	Plasmid pRECH2. N
C 529	16	1.2	3379	22	AAQ66992	Vector pCGMT. Syn	C 602	16	1.2	3784	11	AAQ05326	Sequence encoding
C 530	16	1.2	3383	11	AAQ06309	Sequence of plasmid	C 603	16	1.2	3786	15	AAQ73378	STB5-CRR chimera
C 531	16	1.2	3383	21	AAQ51632	Plasmid pGM678 con	C 604	16	1.2	3789	15	AAQ44362	Sequence of bovine
C 532	16	1.2	3384	21	AAZ29087	CNA of vector pGF	C 605	16	1.2	3789	21	AAZ39782	Plasmid pGB1MN nu
C 533	16	1.2	3400	15	AAQ71366	E.coli/S.cerevisia	C 606	16	1.2	3794	16	AAV00680	Plasmid pTRH46 con
C 534	16	1.2	3403	9	AAAB0956	Plasmid pDS5/RBSII	C 607	16	1.2	3796	21	AAA27831	Vector plasmid pCM
C 535	16	1.2	3414	12	AAQ12785	pDS5/RBSII-2 sequ	C 608	16	1.2	3803	11	AAQ03155	Sequence of the ch
C 536	16	1.2	3414	14	AAQ34609	Expression plasmid	C 609	16	1.2	3819	18	AAZ78825	Kappa light chain
C 537	16	1.2	3415	12	AAQ12784	pDS5/RBSII-1 sequ	C 610	16	1.2	3819	16	AAV39266	Plasmid pRC65 nuc
C 538	16	1.2	3415	14	AAQ34608	Expression plasmid	C 611	16	1.2	3819	20	AAZ22020	Nucleotide sequenc
C 539	16	1.2	3416	9	AAAB0958	Plasmid pDS5/RBSI	C 612	16	1.2	3820	11	AAQ03517	Plasmid pGHI enco
C 540	16	1.2	3416	11	AAQ05302	Sequence of plasmid	C 613	16	1.2	3822	19	AAV13171	Complete DNA sequ
C 541	16	1.2	3416	12	AAQ12783	pDS5/RBSII sequen	C 614	16	1.2	3832	15	AAQ62676	Plasmid pASK60-Str
C 542	16	1.2	3416	14	AAQ34607	Expression plasmid	C 615	16	1.2	3833	10	AAQ91060	Sequence of plasmid
C 543	16	1.2	3418	20	AAZ06440	Plasmid pRZ7075. f	C 616	16	1.2	3840	15	AAQ73376	STB5-CRR chimera
C 544	16	1.2	3419	21	AAAI4902	Nucleotide sequenc	C 617	16	1.2	3840	15	AAQ73377	STB5-CRR chimera
C 545	16	1.2	3420	22	AAH20728	Expression vector	C 618	16	1.2	3850	9	AAAB1634	Sequence of new pl
C 546	16	1.2	3420	22	AAH20728	Expression vector	C 619	16	1.2	3851	18	AAZ79857	Yellowtail tuna DN
C 547	16	1.2	3427	11	AAQ06304	Sequence of plasmid	C 620	16	1.2	3851	18	AAZ90019	Yellowtail fin tu
C 548	16	1.2	3427	11	AAQ06304	Sequence of plasmid	C 621	16	1.2	3853	19	AAV40006	Plasmid pCTM. Chi
C 549	16	1.2	3438	19	AAV39080	S. viridochromogen	C 622	16	1.2	3854	15	AAQ67221	Plasmid pSEC-cyc/c
C 550	16	1.2	3440	11	AAQ06303	Sequence of plasmid	C 623	16	1.2	3858	22	AAQ04947	Plasmid pRK50 used
C 551	16	1.2	3440	12	AAQ13728	pDS5/RBSII. 6xHis	C 624	16	1.2	3861	15	AAV13174	Complete DNA sequ
C 552	16	1.2	3444	21	AAAI5630	Plasmid pGM679 con	C 625	16	1.2	3867	20	AAZ20067	DNA encoding gluta
C 553	16	1.2	3450	22	AAQ66993	Vector pGM6-1b.	C 626	16	1.2	3871	22	AAQ10008	Plasmid pEFM 24 to
C 554	16	1.2	3464	22	AAQ61232	Human DHR gene ex	C 627	16	1.2	3876	21	AAQ98037	Expression vector
C 555	16	1.2	3476	19	AAV48232	Vector plasmid yce	C 628	16	1.2	3876	21	AAQ98038	Expression vector
C 556	16	1.2	3476	19	AAV33299	E.coli-Corynebacte	C 629	16	1.2	3878	14	AAQ40029	Sequence of plasmid
C 557	16	1.2	3480	21	AAQ98011	Expression vector	C 630	16	1.2	3878	16	AAQ76041	Retrovirus vector
C 558	16	1.2	3488	18	AAZ87982	pCALO-2 modular ve	C 631	16	1.2	3878	21	AAQ98015	Plasmid pseip-gpt-
C 559	16	1.2	3519	21	AAQ98035	Expression vector	C 632	16	1.2	3879	21	AAQ98015	Expression vector
C 560	16	1.2	3547	7	AAAB0846	Plasmid sequence e	C 633	16	1.2	3879	21	AAQ98016	Expression vector
C 561	16	1.2	3550	13	AAAB1540	Sequence of pRix-2	C 634	16	1.2	3879	21	AAQ98017	Expression vector
C 562	16	1.2	3550	13	AAQ31623	pRix-2 containing	C 635	16	1.2	3879	21	AAQ98031	Expression vector
C 563	16	1.2	3552	11	AAQ05875	Human pro-growth h	C 636	16	1.2	3879	21	AAQ98036	Expression vector
C 564	16	1.2	3557	7	AAAB0801	Human pro-growth h	C 637	16	1.2	3880	19	AAV39242	Plasmid pCK7-96 nu
C 565	16	1.2	3558	21	AAAZ5006	Nucleotide sequenc	C 638	16	1.2	3881	18	AAZ78801	Kappa light chain
C 566	16	1.2	3585	18	AAZ90399	Plasmid pALV1-781	C 639	16	1.2	3881	20	AAZ21996	Nucleotide sequenc
C 567	16	1.2	3604	21	AAAI5724	Expression vector	C 640	16	1.2	3885	21	AAQ98013	Expression vector
C 568	16	1.2	3618	20	AAZ21890	Nucleotide sequenc	C 641	16	1.2	3888	19	AAV13175	Complete DNA sequ
C 569	16	1.2	3621	7	AAAB0847	Human pre-prolacti	C 642	16	1.2	3888	21	AAV13175	Nucleotide sequenc
C 570	16	1.2	3621	22	AAAF5129	Nucleotide sequenc	C 643	16	1.2	3897	21	AAQ98030	Expression vector
C 571	16	1.2	3623	14	AAQ41297	BMP-2 expression v	C 644	16	1.2	3900	21	AAQ98039	Expression vector
C 572	16	1.2	3628	7	AAAB0848	Plasmid sequence e	C 645	16	1.2	3903	22	AAZ60708	Feline IL-16 encod
C 573	16	1.2	3632	13	AAQ02725	PATRXA/EK/TLII-de	C 646	16	1.2	3907	20	AAV13280	Enterococcus faeca
C 574	16	1.2	3632	13	AAQ027476	pATRXA/EK/TLII-del	C 647	16	1.2	3907	22	AAZ13152	Escherichia coli T
C 575	16	1.2	3632	15	AAQ56898	E.coli thio-redoxin	C 648	16	1.2	3912	22	AAZ13152	Expression vector
C 576	16	1.2	3632	15	AAQ44670	Thio-redoxin-Interi	C 649	16	1.2	3912	21	AAQ98018	Expression vector
C 577	16	1.2	3632	16	AAQ93133	Vector for thio-red	C 650	16	1.2	3919	21	AAQ98028	Expression vector
C 578	16	1.2	3632	16	AAQ93133	Thio-redoxin/des-pr	C 651	16	1.2	3926	21	AAQ03736	Sequence of plasmid
C 579	16	1.2	3634	22	AAAF81730	Human protease and	C 652	16	1.2	3926	22	AAQ10016	Plasmid pSFM 9 to
C 580	16	1.2	3648	22	AAAD65201	Pestivirus NS5A f	C 653	16	1.2	3927	21	AAQ98026	Expression vector
C 581	16	1.2	3661	15	AAQ62675	Plasmid pASK46 for	C 654	16	1.2	3927	21	AAQ98032	Expression vector
C 582	16	1.2	3681	12	AAQ13578	Plasmid pKSELS. S	C 655	16	1.2	3932	22	AAQ98032	Mus musculus goose
C 583	16	1.2	3698	18	AAZ70518	Sequence of vector	C 656	16	1.2	3932	21	AAQ98093	Expression vector
C 584	16	1.2	3698	20	AAQ06038	Vector pGPE. Ratt	C 657	16	1.2	3936	21	AAQ98023	Expression vector
C 585	16	1.2	3699	13	AAQ23445	Plasmid pGPE for c	C 658	16	1.2	3938	20	AAQ98023	Plasmid pPA6-Kanx
C 586	16	1.2	3699	14	AAQ44179	Heavy chain minilo	C 659	16	1.2	3944	13	AAQ29750	PCMX vector. Synt
C 587	16	1.2	3699	17	AAZ37237	Plasmid pBSGFP exp	C 660	16	1.2	3945	21	AAQ98033	Expression vector
C 588	16	1.2	3699	18	AAV14340	Plasmid pGPE. Syn	C 661	16	1.2	3950	22	AAQ98033	Plasmid pSB-92 nuc
C 589	16	1.2	3699	18	AAV12534	Plasmid pGPE #2.	C 662	16	1.2	3954	21	AAQ98034	Nucleotide sequenc
C 590	16	1.2	3699	18	AAV02619	Plasmid pGPE #2.	C 663	16	1.2	3954	22	AAQ98034	Plasmid pGHR1-297W
C 591	16	1.2	3705	22	AAQ83391	Unique BglII site	C 664	16	1.2	3955	19	AAV13173	Complete DNA sequ
C 592	16	1.2	3712	16	AAQ97490	Plasmid pTRAI. N	C 665	16	1.2	3956	19	AAV64258	Plasmid pPR13/14 D
C 593	16	1.2	3723	22	AAQ10000	Plasmid pCMH42 co	C 666	16	1.2	3957	21	AAQ98034	Expression vector
C 594	16	1.2	3728	22	AAQ04929	Retroviral control	C 667	16	1.2	3968	22	AAQ98034	pHS70-1MCS const
C 595	16	1.2	3735	12	AAQ14452	Rat prolactin gene	C 668	16	1.2	3976	13	AAQ25119	Expression vector

C 669	16	1.2	3976	22	AA086256	742	16	1.2	4173	19	AAV32653	Plasmid pGFP-TT re
C 670	16	1.2	3977	14	AA049223	743	16	1.2	4176	10	AAV90709	Sequence of plasmid
C 671	16	1.2	3977	15	AA087695	744	16	1.2	4184	21	AA16546	Human secreted pro
C 672	16	1.2	3977	15	AA087696	745	16	1.2	4189	11	AA05397	Secretion Vector p
C 673	16	1.2	3977	15	AA087700	746	16	1.2	4189	11	AA05397	Plasmid pMW12 con
C 674	16	1.2	3977	15	AA087701	747	16	1.2	4189	11	AA05397	H. ghiliani1/B. me
C 675	16	1.2	3977	15	AA087702	748	16	1.2	4201	21	AA097949	Human cancer assoc
C 676	16	1.2	3977	15	AA087703	749	16	1.2	4201	21	AA097949	Human colon cancer
C 677	16	1.2	3977	15	AA087704	750	16	1.2	4201	22	AA033277	Feline IL-12p35 en
C 678	16	1.2	3977	15	AA087705	751	16	1.2	4201	22	AA033277	Sequence of recomb
C 679	16	1.2	3977	15	AA087706	752	16	1.2	4201	22	AA033277	Vector pEP1ink6 D
C 680	16	1.2	3977	15	AA087707	753	16	1.2	4214	21	AA062652	Nucleotide sequenc
C 681	16	1.2	3977	15	AA087708	754	16	1.2	4224	21	AA092038	Sequence of recomb
C 682	16	1.2	3977	15	AA087709	755	16	1.2	4224	22	AA055132	Nucleotide sequenc
C 683	16	1.2	3977	15	AA087710	756	16	1.2	4229	22	AA025896	Genetic informatio
C 684	16	1.2	3977	15	AA087711	757	16	1.2	4229	19	AA063466	Plasmid pCMTIE. C
C 685	16	1.2	3977	15	AA087712	758	16	1.2	4257	20	AA022251	Nucleotide sequenc
C 686	16	1.2	3977	15	AA087684	759	16	1.2	4259	21	AA092002	PCR5-EGFP report
C 687	16	1.2	3977	15	AA087687	760	16	1.2	4274	14	AA054144	Sequence of plasmid
C 688	16	1.2	3977	15	AA087688	761	16	1.2	4276	22	AA06790	Plasmid pMW11. C
C 689	16	1.2	3977	15	AA087689	762	16	1.2	4277	14	AA040280	Sequence of clone
C 690	16	1.2	3977	15	AA087690	763	16	1.2	4277	21	AA089874	Plasmid pS29pt-P2.
C 691	16	1.2	3977	15	AA087690	764	16	1.2	4282	22	AA083668	PCMV-IT nucleic ac
C 692	16	1.2	3977	15	AA087684	765	16	1.2	4283	18	AA086449	DNA encoding hgh.
C 693	16	1.2	3977	15	AA087685	766	16	1.2	4283	19	AA086449	Epidermal growth f
C 694	16	1.2	3977	15	AA073431	767	16	1.2	4283	21	AA063227	Plasmid pMG1630 n
C 695	16	1.2	3977	15	AA087697	768	16	1.2	4285	21	AA075780	Human ORF135
C 696	16	1.2	3977	15	AA087691	769	16	1.2	4285	21	AA075780	Vector MSX3-DRT10-
C 697	16	1.2	3977	15	AA087692	770	16	1.2	4299	21	AA027980	Vector ORF3-DBT10-
C 698	16	1.2	3977	15	AA087692	771	16	1.2	4300	21	AA029135	PCR5-EGFP const
C 699	16	1.2	3977	15	AA087693	772	16	1.2	4308	22	AA02649	DN305 plasmid for
C 700	16	1.2	3977	15	AA087699	773	16	1.2	4330	15	AA026524	Plasmid pPUL2 enco
C 701	16	1.2	3977	15	AA087694	774	16	1.2	4341	15	AA062391	Vector pVAC1. Syn
C 702	16	1.2	3979	11	AA006512	775	16	1.2	4343	21	AA075007	Nucleotide sequenc
C 703	16	1.2	3979	11	AA051643	776	16	1.2	4344	21	AA053848	Vector pTGF67. S
C 704	16	1.2	3983	22	AA010265	777	16	1.2	4357	22	AA031389	Expression vector
C 705	16	1.2	3984	16	AA007310	778	16	1.2	4366	16	AA085424	Plasmid pTREC3-P28
C 706	16	1.2	3988	22	AA087639	779	16	1.2	4368	21	AA098020	Expression vector
C 707	16	1.2	3988	22	AA087639	780	16	1.2	4369	21	AA098020	Partial sequenc o
C 708	16	1.2	3989	21	AA086025	781	16	1.2	4377	16	AA085421	Plasmid pTREC1-P28
C 709	16	1.2	3991	10	AA091061	782	16	1.2	4380	16	AA098022	Expression vector
C 710	16	1.2	3992	21	AA097948	783	16	1.2	4382	22	AA055131	Nucleotide sequenc
C 711	16	1.2	3999	20	AA002663	784	16	1.2	4386	22	AA010014	Plasmid pSM 5 to
C 712	16	1.2	4003	11	AA006305	785	16	1.2	4390	21	AA056322	CRS-P30-CRS fusion
C 713	16	1.2	4004	11	AA006306	786	16	1.2	4396	21	AA056322	atb cloning vecto
C 714	16	1.2	4005	14	AA040418	787	16	1.2	4403	17	AA048899	Eradixoxina-trip f
C 715	16	1.2	4009	19	AA013172	788	16	1.2	4421	17	AA058319	DNA sequence of he
C 716	16	1.2	4009	19	AA000679	789	16	1.2	4432	15	AA074206	Expression vector
C 717	16	1.2	4019	20	AA002663	790	16	1.2	4432	16	AA071113	Expression vector
C 718	16	1.2	4026	19	AAV40007	791	16	1.2	4432	22	AA009426	Vaccine vector VJ
C 719	16	1.2	4032	19	AAV23237	792	16	1.2	4433	22	AA090612	Arabidopsis thaila
C 720	16	1.2	4045	15	AA070942	793	16	1.2	4433	18	AA095023	Plasmid pCMT1.8 co
C 721	16	1.2	4045	22	AA091149	794	16	1.2	4443	8	AA086663	Murine anti-botuli
C 722	16	1.2	4059	20	AA07338	795	16	1.2	4443	22	AA070627	Encodes amplicillin
C 723	16	1.2	4059	16	AA001001	796	16	1.2	4450	22	AA050237	Plasmid pAN441 for
C 724	16	1.2	4059	17	AA01899	797	16	1.2	4451	21	AA025637	CRS-P33-CRS fusion
C 725	16	1.2	4059	17	AA01899	798	16	1.2	4451	21	AA025637	Human cyclin DI-hu
C 726	16	1.2	4059	20	AA012239	799	16	1.2	4453	19	AA069890	Human cyclin DI/cy
C 727	16	1.2	4059	20	AA012239	800	16	1.2	4454	14	AA054118	Plasmid pSV16B5, f
C 728	16	1.2	4059	20	AA012239	801	16	1.2	4454	18	AA066462	Plasmid pCMT1.8 co
C 729	16	1.2	4059	20	AA012239	802	16	1.2	4454	22	AA066462	Vector pGX104 DNA
C 730	16	1.2	4068	9	AA009957	803	16	1.2	4464	22	AA066462	Plasmid pGHRH-29YA
C 731	16	1.2	4074	21	AA0248267	804	16	1.2	4466	15	AA080286	Plasmid pPR104-4.
C 732	16	1.2	4088	19	AA064255	805	16	1.2	4466	15	AA080286	pPR104-4 sequenc.
C 733	16	1.2	4100	21	AA050629	806	16	1.2	4477	13	AA021962	Sequence of pHCV-3
C 734	16	1.2	4102	19	AA064257	807	16	1.2	4477	13	AA021962	Sequence of pHCV-3
C 735	16	1.2	4114	16	AA090777	808	16	1.2	4481	14	AA031677	recombin
C 736	16	1.2	4118	18	AA069188	809	16	1.2	4481	14	AA031677	recombin
C 737	16	1.2	4118	18	AA069189	810	16	1.2	4481	14	AA031677	recombin
C 738	16	1.2	4145	14	AA040279	811	16	1.2	4481	22	AA031677	recombin
C 739	16	1.2	4145	21	AA089673	812	16	1.2	4482	15	AA022652	recombin
C 740	16	1.2	4163	13	AA032349	813	16	1.2	4482	15	AA022652	recombin
C 741	16	1.2	4172	20	AA019901	814	16	1.2	4496	22	AA09266	recombin

815	16	1.2	4503	19	AAV13170	Complete DNA sequ
816	16	1.2	4515	14	AAO46823	PLIS-SF vector
817	16	1.2	4522	22	AAE28076	Feline IL-12p40 en
818	16	1.2	4522	22	AAE28076	Nucleotide seqenc
819	16	1.2	4525	20	AAE69746	Nucleotide seqenc
820	16	1.2	4539	16	AAO87347	PIWV plasmid used
821	16	1.2	4539	19	AAV37292	Human cyclin D1-hu
822	16	1.2	4540	18	AAE69891	Human cyclin D1/cy
823	16	1.2	4540	19	AAV54119	atrr reading frame
824	16	1.2	4554	21	AAE55541	Vector plasmid pW6
825	16	1.2	4570	18	AAE90491	Single chain Fv an
826	16	1.2	4574	19	AAV32664	Plasmid pGL25 stre
827	16	1.2	4583	10	AAV90183	DNA sequence of p1
828	16	1.2	4583	12	AAO15174	Plasmid pAMVBT4
829	16	1.2	4583	18	AAE59737	Plasmid pPRK7/8 DNA
830	16	1.2	4583	19	AAE64256	Sequence of plasmid
831	16	1.2	4583	19	AAE64256	DNA sequence of th
832	16	1.2	4583	21	AAV3441	Plasmid pSUN387
833	16	1.2	4603	11	AAO40410	Sequence of new re
834	16	1.2	4608	10	AAV90288	Nucleotide sequenc
835	16	1.2	4608	10	AAV90288	Plasmid pUCMS-6h73
836	16	1.2	4610	21	AAE59005	Plasmid pROCOS4/7
837	16	1.2	4613	22	AAE59005	Vector pROCOS4/7
838	16	1.2	4613	22	AAE59005	Rat tripeptidylpep
839	16	1.2	4617	17	AAV5803	Vector pGNS DNA se
840	16	1.2	4617	17	AAV5803	Human cyclin D1-hu
841	16	1.2	4618	22	AAE39638	Human cyclin D1/cy
842	16	1.2	4621	18	AAE69889	Plasmid pBS13
843	16	1.2	4621	19	AAV54117	pHSP-OHDXS/BI pla
844	16	1.2	4621	19	AAV54117	Sequence of plasmid
845	16	1.2	4626	18	AAE9498	DHFR-APP fusion pr
846	16	1.2	4626	22	AAO3988	Vector pGK52 DNA s
847	16	1.2	4641	14	AAO40294	Recombinant plasmid
848	16	1.2	4643	13	AAO25112	Alpha-glucosidase
849	16	1.2	4643	13	AAO25112	Sequence of recomb
850	16	1.2	4644	21	AAE53850	DHFR-APP fusion pr
851	16	1.2	4644	21	AAE53850	Vector pGK52 DNA s
852	16	1.2	4648	12	AAO12756	Recombinant plasmid
853	16	1.2	4651	10	AAE25108	Alpha-glucosidase
854	16	1.2	4651	10	AAE25108	Sequence of recomb
855	16	1.2	4657	19	AAV5804	DHFR-APP fusion pr
856	16	1.2	4657	19	AAV5804	Vector pROCOS5/3
857	16	1.2	4660	10	AAV90266	Plasmid pPRG1F2
858	16	1.2	4668	11	AAO4519	Recombinant plasmid
859	16	1.2	4670	13	AAO25107	DHFR-APP fusion pr
860	16	1.2	4673	13	AAO25107	Sequence of recomb
861	16	1.2	4688	10	AAE25107	Plasmid pGK400 for
862	16	1.2	4689	21	AAE24986	RBI, PDI coexpress
863	16	1.2	4690	13	AAO29910	Nucleotide sequenc
864	16	1.2	4691	16	AAO32546	Plasmid pRT-L2
865	16	1.2	4694	22	AAE55225	Vector pG218 DNA s
866	16	1.2	4701	22	AAE55225	Murine IAD alpha c
867	16	1.2	4701	22	AAE55225	Plasmid pGFR2-15 c
868	16	1.2	4713	19	AAV12068	Plasmid pSG1-12 of
869	16	1.2	4715	11	AAO50284	Gamma heavy chain
870	16	1.2	4715	11	AAO50284	Plasmid pCG7-96 nu
871	16	1.2	4716	11	AAO5376	Nucleotide sequenc
872	16	1.2	4723	18	AAE78802	Nucleotide sequenc
873	16	1.2	4723	19	AAV39243	Murine IAD beta ch
874	16	1.2	4723	20	AAE21997	Human ORFX ORF1479
875	16	1.2	4723	22	AAE55224	PT7-TACS vector DN
876	16	1.2	4724	19	AAV12068	Recombinant plasmid
877	16	1.2	4724	21	AAE75924	Plasmid pRK5.1 for
878	16	1.2	4727	22	AAO39265	Nucleotide sequenc
879	16	1.2	4732	14	AAO51730	Nucleotide sequenc
880	16	1.2	4732	14	AAO51730	Vector pGK22 DNA s
881	16	1.2	4732	22	AAE55224	DHFR-APP fusion pr
882	16	1.2	4756	22	AAE55224	Portion of plasmid
883	16	1.2	4760	11	AAO52099	Plasmid pBSOLEX1
884	16	1.2	4760	11	AAO52099	Plasmid pBLAKI sta
885	16	1.2	4762	9	AAE81675	
886	16	1.2	4768	22	AAE50243	
887	16	1.2	4774	9	AAE81424	
888	16	1.2	4775	21	AAE56319	CRS-P29-CRS fusion
889	16	1.2	4776	20	AAE77617	Expression constru
890	16	1.2	4776	20	AAE77617	Expression constru
891	16	1.2	4779	9	AAE82213	Plasmid pGFI enco
892	16	1.2	4789	16	AAE84696	Plasmid pGFI enco
893	16	1.2	4792	16	AAE84696	Plasmid pGFI enco
894	16	1.2	4816	22	AAO27799	PTCHS1B vector co
895	16	1.2	4816	22	AAO27799	Plasmid pPRAT-delta
896	16	1.2	4819	15	AAE44360	Plasmid pPRAT-delta
897	16	1.2	4819	18	AAE89768	Plasmid pPRAT-delta
898	16	1.2	4819	18	AAE89768	Plasmid pPRAT-delta
899	16	1.2	4819	18	AAE89768	Plasmid pPRAT-delta
900	16	1.2	4824	17	AAE56594	Shortened C1 gene
901	16	1.2	4824	17	AAE56594	Nucleotide sequenc
902	16	1.2	4824	22	AAE15634	Plasmid pBFM 25 ex
903	16	1.2	4840	21	AAE88447	Plasmid pB0475 enc
904	16	1.2	4846	22	AAE83146	Complete sequenc
905	16	1.2	4846	22	AAE83146	P. pyralis pretluc
906	16	1.2	4846	21	AAE29133	pN-kappaB-SEAP co
907	16	1.2	4847	21	AAE29133	pN-kappaB-SEAP co
908	16	1.2	4867	18	AAE47448	Human growth hormo
909	16	1.2	4867	18	AAE47448	Plasmid pB0475 inc
910	16	1.2	4867	11	AAO46571	Human growth hormo
911	16	1.2	4867	20	AAE87976	Plasmid pB0475 con
912	16	1.2	4883	13	AAE25709	pDEL10. Synthetic
913	16	1.2	4883	13	AAE25709	pDEL10. Synthetic
914	16	1.2	4883	16	AAO51192	Plasmid pKSI nucle
915	16	1.2	4896	16	AAO8976	PT256 contg. P355
916	16	1.2	4896	20	AAE15631	Nucleotide sequenc
917	16	1.2	4905	21	AAE00395	Mammalian expressi
918	16	1.2	4906	10	AAE95081	Plasmid pPR709 co
919	16	1.2	4909	16	AAE18924	Plasmid pA1261. S
920	16	1.2	4910	17	AAE44357	Plasmid pPRATPE
921	16	1.2	4910	17	AAE15287	pPR78/52 hybrid a
922	16	1.2	4910	18	AAE89761	Plasmid pPRATPE
923	16	1.2	4910	18	AAE89761	Plasmid pPRATPE
924	16	1.2	4910	18	AAE89761	Plasmid pPRATPE
925	16	1.2	4916	13	AAE22882	Synthetic human gr
926	16	1.2	4924	21	AAE34937	Retrovirus vector
927	16	1.2	4924	22	AAE30946	Vector used in inv
928	16	1.2	4926	19	AAE39291	Plasmid pPRATPE
929	16	1.2	4926	20	AAE22045	Nucleotide sequenc
930	16	1.2	4930	21	AAE75000	Nucleotide sequenc
931	16	1.2	4930	21	AAE75000	Nucleotide sequenc
932	16	1.2	4930	21	AAE75000	Nucleotide sequenc
933	16	1.2	4930	21	AAE75000	Nucleotide sequenc
934	16	1.2	4930	21	AAE75000	Nucleotide sequenc
935	16	1.2	4930	21	AAE75000	Nucleotide sequenc
936	16	1.2	4930	21	AAE75000	Nucleotide sequenc
937	16	1.2	4930	21	AAE75000	Nucleotide sequenc
938	16	1.2	4930	21	AAE75000	Nucleotide sequenc
939	16	1.2	4930	21	AAE75000	Nucleotide sequenc
940	16	1.2	4930	21	AAE75000	Nucleotide sequenc
941	16	1.2	4930	21	AAE75000	Nucleotide sequenc
942	16	1.2	4930	21	AAE75000	Nucleotide sequenc
943	16	1.2	4930	21	AAE75000	Nucleotide sequenc
944	16	1.2	4930	21	AAE75000	Nucleotide sequenc
945	16	1.2	4930	21	AAE75000	Nucleotide sequenc
946	16	1.2	4930	21	AAE75000	Nucleotide sequenc
947	16	1.2	4930	21	AAE75000	Nucleotide sequenc
948	16	1.2	4930	21	AAE75000	Nucleotide sequenc
949	16	1.2	4930	21	AAE75000	Nucleotide sequenc
950	16	1.2	4930	21	AAE75000	Nucleotide sequenc
951	16	1.2	4930	21	AAE75000	Nucleotide sequenc
952	16	1.2	4930	21	AAE75000	Nucleotide sequenc
953	16	1.2	4930	21	AAE75000	Nucleotide sequenc
954	16	1.2	4930	21	AAE75000	Nucleotide sequenc
955	16	1.2	4930	21	AAE75000	Nucleotide sequenc
956	16	1.2	4930	21	AAE75000	Nucleotide sequenc
957	16	1.2	4930	21	AAE75000	Nucleotide sequenc
958	16	1.2	4930	21	AAE75000	Nucleotide sequenc
959	16	1.2	4930	21	AAE75000	Nucleotide sequenc
960	16	1.2	4930	21	AAE75000	Nucleotide sequenc
961	16	1.2	4930	21	AAE75000	Nucleotide sequenc
962	16	1.2	4930	21	AAE75000	Nucleotide sequenc
963	16	1.2	4930	21	AAE75000	Nucleotide sequenc
964	16	1.2	4930	21	AAE75000	Nucleotide sequenc
965	16	1.2	4930	21	AAE75000	Nucleotide sequenc
966	16	1.2	4930	21	AAE75000	Nucleotide sequenc
967	16	1.2	4930	21	AAE75000	Nucleotide sequenc
968	16	1.2	4930	21	AAE75000	Nucleotide sequenc
969	16	1.2	4930	21	AAE75000	Nucleotide sequenc
970	16	1.2	4930	21	AAE75000	Nucleotide sequenc
971	16	1.2	4930	21	AAE75000	Nucleotide sequenc
972	16	1.2	4930	21	AAE75000	Nucleotide sequenc
973	16	1.2	4930	21	AAE75000	Nucleotide sequenc
974	16	1.2	4930	21	AAE75000	Nucleotide sequenc
975	16	1.2	4930	21	AAE75000	Nucleotide sequenc
976	16	1.2	4930	21	AAE75000	Nucleotide sequenc
977	16	1.2	4930	21	AAE75000	Nucleotide sequenc
978	16	1.2	4930	21	AAE75000	Nucleotide sequenc
979	16	1.2	4930	21	AAE75000	Nucleotide sequenc
980	16	1.2	4930	21	AAE75000	Nucleotide sequenc
981	16	1.2	4930	21	AAE75000	Nucleotide sequenc
982	16	1.2	4930	21	AAE75000	Nucleotide sequenc
983	16	1.2	4930	21	AAE75000	Nucleotide sequenc
984	16	1.2	4930	21	AAE75000	Nucleotide sequenc
985	16	1.2	4930	21	AAE75000	Nucleotide sequenc
986	16	1.2	4930	21	AAE75000	Nucleotide sequenc
987	16	1.2	4930	21	AAE75000	Nucleotide sequenc
988	16	1.2	4930	21	AAE75000	Nucleotide sequenc
989	16	1.2	4930	21	AAE75000	Nucleotide sequenc
990	16	1.2	4930	21	AAE75000	Nucleotide sequenc
991	16	1.2	4930	21	AAE75000	Nucleotide sequenc
992	16	1.2	4930	21	AAE75000	Nucleotide sequenc
993	16	1.2	4930	21	AAE75000	Nucleotide sequenc
994	16	1.2	4930	21	AAE75000	Nucleotide sequenc
995	16	1.2	4930	21	AAE75000	Nucleotide sequenc
996	16	1.2	4930	21	AAE75000	Nucleotide sequenc
997	16	1.2	4930	21	AAE75000	Nucleotide sequenc
998	16	1.2	4930	21	AAE75000	

C 961	16	1.2	5015	18	AA147449	Human somatogenic
C 962	16	1.2	5015	20	AAx87977	Human soluble grow
C 963	16	1.2	5015	20	AAV62766	Plasmid pJ1446 con
C 964	16	1.2	5015	21	AA288448	Plasmid pJ1446 enc
C 965	16	1.2	5021	21	AAx07775	DNA sequence of pl
C 966	16	1.2	5027	20	AAx78865	Human tissue facto
C 967	16	1.2	5030	18	AA775702	pgreenlantern vect
C 968	16	1.2	5030	20	AAx08453	pgreenlantern-1 ve
C 969	16	1.2	5033	15	AAQ74695	pPH1406 vector co
C 970	16	1.2	5042	16	AAQ84694	Plasmid glucanyla
C 971	16	1.2	5051	22	AAQ84694	Nucleotide sequenc
C 972	16	1.2	5056	10	AA907095	Plasmid pCB109
C 973	16	1.2	5060	18	AA778171	Plasmid pK122-8
C 974	16	1.2	5062	18	AA778170	Plasmid pSE-2 enco
C 975	16	1.2	5064	21	AAV98021	Expression vector
C 976	16	1.2	5067	19	AAV18740	Complete DNA seque
C 977	16	1.2	5069	20	AAx78871	Human tissue facto
C 978	16	1.2	5096	16	AAQ83824	Plasmid pK122-8
C 979	16	1.2	5098	15	AAQ84356	Plasmid pCat70PE
C 980	16	1.2	5098	18	AA789760	Plasmid pCat70PE
C 981	16	1.2	5098	18	AA788225	Plasmid pCat70PE
C 982	16	1.2	5098	18	AAV84568	Plasmid pCat70PE e
C 983	16	1.2	5098	18	AAV56584	Plasmid pCat70PE g
C 984	16	1.2	5099	13	AAQ24687	DNA encoding B-HDA
C 985	16	1.2	5101	21	AA10799	Nucleotide sequenc
C 986	16	1.2	5104	22	AAQ07798	ptpCH3B vector co
C 987	16	1.2	5109	17	AA772654	Kat brain homology
C 988	16	1.2	5109	19	AAV13846	Complete DNA seque
C 989	16	1.2	5110	16	AAQ83833	Plasmid PCM-B64
C 990	16	1.2	5110	16	AAQ83834	Plasmid PCM-p1208
C 991	16	1.2	5110	16	AAQ83835	Plasmid PCM-p1031
C 992	16	1.2	5110	16	AAQ83836	Plasmid PCM-T221
C 993	16	1.2	5110	16	AAQ83825	Plasmid PCM-X#-S
C 994	16	1.2	5110	16	AAQ83825	Plasmid PCM-X#-S
C 995	16	1.2	5110	16	AAQ83827	Plasmid PCM-T297
C 996	16	1.2	5110	16	AAQ83828	Plasmid PCM-p1160
C 997	16	1.2	5110	16	AAQ83829	Plasmid PCM-T270
C 998	16	1.2	5110	16	AAQ83830	Plasmid PCM-p1087
C 999	16	1.2	5110	16	AAQ83831	Plasmid PCM-p1198
C1000	16	1.2	5110	16	AAQ83832	Plasmid PCM-T286

ALIGNMENTS

RESULT 1

AAS00045 standard; DNA: 1362 BP.

AC AAS00045;

DT 14-MAY-2001 (first entry)

XX Human DNA encoding C2GnT3.

XX Human: C2GnT3; Thymus-related disorder; cancer; tumour; adenoma;

XX UDP-GlcNAc: Galbeta1,3galNAc alpha-R beta1,6GalNAc transferase; sarcoma;

XX malignant melanoma; breast cancer; cervical cancer; hypodactylity;

XX hypodactylity; atrophy; thymus enlargement; autoimmunity; arthritis;

XX Leukemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;

XX acquired immunodeficiency syndrome; sepsis; wound healing; infection; ds.

XX Homo sapiens.

XX Key

XX CDS

FT primer_bind

FT primer_bind

FT primer_bind

FT primer_bind

FT primer_bind

/label= "Binding site for PCR primer TSHC121"

W0200114535-A2.
01-MAR-2001.
24-AUG-2000; 2000WO-DK00469.
24-AUG-1999; 99US-0150488.
(SCHW)/ SCHWIENTEK T.
(CLAU)/ CLAUSEN H.
Schwientek T, Clausen H;
WPI: 2001-226615/23.
P-PSDB; AAU00037.
New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
thymus-related disorders, cancers, tumours, immunosuppression

Claim 5; Fig 1; 97pb; English.

The sequence encodes Human UDP-N-acetyl-glucosamine:Galactose-beta1,
3-N-acetylglucosamine-alpha-R beta1-6-N-acetylglucosaminyltransferase
(UDP-GlcNAc: Galbeta1,3galNAc alpha-R beta1,6GalNAc transferase or
C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
preparation of compositions for treating a conditions mediated by C2GnT3,
particularly a thymus-related disorder. C2GnT3, nucleic acids
encoding it and antibodies against it may also be used for in vitro
purposes related to scientific research, DNA synthesis and manufacture of
vectors, in the prognostic and diagnostic evaluation of conditions
associated with altered expression or activity of C2GnT3 or conditions
requiring modulation of C2GnT3, as well as in monitoring conditions by
detecting and localising the DNA and protein. Disorders such as tumours
(e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
breast or cervix), hypodactylity, hyperactivity, atrophy, enlargement of
thymus, autoimmunity, arthritis, leukemia, lymphoma, immunosuppression,
acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
sepsis, wound healing, acute and chronic infection, cell-mediated or
humour immunity, or Th1/Th2 imbalance, may be treated using these protein
or nucleic acid. The antibodies may be used to screen potential
therapeutic compounds to determine their effects on a conditions such as
thymus-related disorder or cancer, to determine the level of C2GnT3
expression in cells genetically engineered to produce C2GnT3, or to
detect and quantify polypeptides in a sample to determine their role in a
particular cellular events or pathological states and to diagnose and
treat such pathological states.

Sequence 1362 BP; 410 A; 260 C; 279 G; 413 T; 0 other;

Query Match 100.0%; Score 1362; DB 22; Length 1362;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagatattcaaatgtattttaacatcaccctcagagaaagtttcatctgttt 60
 Db 1 atgaagatattcaaatgtattttaacatcaccctcagagaaagtttcatctgttt 60
 QY 61 ttaaccctatgctcctctcttctgtttaagcttcaaatgtagaagcctttccgcaa 120
 Db 61 ttaaccctatgctcctctcttctgtttaagcttcaaatgtagaagcctttccgcaa 120
 QY 121 aaagacattacttggttgagtaactccctaagtaactcgcctttgttaagaacagatac 180
 Db 121 aaagacattacttggttgagtaactccctaagtaactcgcctttgttaagaacagatac 180
 QY 181 actcatgttaagatgaagtaagatgaagtaagtaagtaagtaagtaagtaagtaagtaag 240
 Db 181 actcatgttaagatgaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 240

```

QY 241 ccttggaaattgaaagagctggaataagaagaaggagacatctactgtgagat 300
    |||||
Db 241 ccttggaaattgaaagagctggaataagaagaaggagacatctactgtgagat 300

QY 301 gatgatgtgtggcaatgagcagtgattgtgacattatcagaacttaagagctatgct 360
    |||||
Db 301 gatgatgtgtggcaatgagcagtgattgtgacattatcagaacttaagagctatgct 360

QY 361 caaaagctgtctcaagagagagaaagcttcccaatgagcattcttgtgtccac 420
    |||||
Db 361 caaaagctgtctcaagagagagaaagcttcccaatgagcattcttgtgtccac 420

QY 421 aaagatgcaattatgttgaagagctatccatgctatataacaccagacaatatattac 480
    |||||
Db 421 aaagatgcaattatgttgaagagctatccatgctatataacaccagacaatatattac 480

QY 481 tgcatacatatgatgtgaagcagctgtatccttcaagtgccatgagcaatattagct 540
    |||||
Db 481 tgcatacatatgatgtgaagcagctgtatccttcaagtgccatgagcaatattagct 540

QY 541 aagtgtcttcccaatatttcatgtgtccaaattagaagctgtggaataatgccacat 600
    |||||
Db 541 aagtgtcttcccaatatttcatgtgtccaaattagaagctgtggaataatgccacat 600

QY 601 tccagactccaagctgaatttaaatgtgtgagacccctcctgaagcttccatccagt 660
    |||||
Db 601 tccagactccaagctgaatttaaatgtgtgagacccctcctgaagcttccatccagt 660

QY 661 aaatatgtatccaactgtgtggcaagatcttccctgagatccaatttggatggct 720
    |||||
Db 661 aaatatgtatccaactgtgtggcaagatcttccctgagatccaatttggatggct 720

QY 721 tcagagtgtgaaaaaactcaatgagcaaatatgttggagaagctgaaccccaacagt 780
    |||||
Db 721 tcagagtgtgaaaaaactcaatgagcaaatatgttggagaagctgaaccccaacagt 780

QY 781 aaattggaagatcatctacatcagactgaacgagtgcttgaataatgtgagag 840
    |||||
Db 781 aaattggaagatcatctacatcagactgaacgagtgcttgaataatgtgagag 840

QY 841 ctaccaaataagacaacacatcccaaggaagacccccccataacattcagatattgt 900
    |||||
Db 841 ctaccaaataagacaacacatcccaaggaagacccccccataacattcagatattgt 900

QY 901 ggcagctctatttgtttaaagtcagacattgttaaatatttccaacatccatcc 960
    |||||
Db 901 ggcagctctatttgtttaaagtcagacattgttaaatatttccaacatccatcc 960

QY 961 gttaaaagacttttgcctgtgtctaaagacacatactctctgatgagcatttggct 1020
    |||||
Db 961 gttaaaagacttttgcctgtgtctaaagacacatactctctgatgagcatttggct 1020

QY 1021 accctgttggtgtccagaataacctgggagatttccagatcagcccgagtggtct 1080
    |||||
Db 1021 accctgttggtgtccagaataacctgggagatttccagatcagcccgagtggtct 1080

QY 1081 gatctgagagtaagactcgcctgtgcaagtggaaattactactaagagcttttctatccc 1140
    |||||
Db 1081 gatctgagagtaagactcgcctgtgcaagtggaaattactactaagagcttttctatccc 1140

QY 1141 agttgtactgagatcccttccagagctgtgtattatgagagctgagaaattaaagttg 1200
    |||||
Db 1141 agttgtactgagatcccttccagagctgtgtattatgagagctgagaaattaaagttg 1200

QY 1201 ctatcaaaagatgacattgtgtgcaataaatattgattcctaagtgagcccatcttg 1260
    |||||
Db 1201 ctatcaaaagatgacattgtgtgcaataaatattgattcctaagtgagcccatcttg 1260

QY 1261 attaatgtcttgagcaaaaagcttgaagaacagagagagactggatcactttgcccctca 1320
    |||||
Db 1261 attaatgtcttgagcaaaaagcttgaagaacagagagagactggatcactttgcccctca 1320

QY 1321 gaaagttattatgagtagaatactcactaccatcatgaa 1362

```

```

Db 1321 gaaagttattatgagtagaatactcactaccatcatgaa 1362

RESULT 2
AAH98678
ID AAH98678 standard; cDNA; 781 BP.
AC AAH98678;
XX
XX
XX 12-OCT-2001 (first entry)
XX
XX Human EST-derived coding sequence SEQ ID NO: 535.
DE
XX
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KM biodiversity; gene therapy; nutrition; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200154477-A2.
XX
XX 02-AUG-2001.
PD
XX
XX 25-JAN-2001; 2001WO-US02687.
PF
XX
XX 25-JAN-2000; 2000US-0491404.
PR
XX 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
XX (HXYE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
XX DR MPI: 2001-476164/51.
XX
XX P-PSDB; AMM24019.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
XX Claim 1: Page 549; 1275pp; English.
XX
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
XX Sequence 781 BP; 218 A; 175 C; 174 G; 214 T; 0 other;
SQ

Query Match 25.6%; Score 349; DB 22; Length 781;
Best Local Similarity 100.0%; Pred. No. 1,3e-157;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY 181 actcatgttaagagatgaagtcagatgaagtaactgttcgggtatcatatgaacaaggag 240
Db 410 actcatgttaagagatgaagtcagatgaagtaactgttcgggtatcatatgaacaaggag 469
QY 241 cctttggaatgtgaagaagctctctggaataagaagaaggacatcatgtgaggagat 300
Db 470 cctttggaatgtgaagaagctctctggaataagaagaaggacatcatgtgaggagat 529
QY 301 gatgatgtgtggaatgacacagatgtgtgacattatcacagactctaa 349
Db 530 gatgatgtgtggaatgacacagatgtgtgacattatcacagactctaa 578

RESULT 3
AACI9209
ID AACI9209 standard; cDNA; 186 BP.
XX
XX AACI9209;
AC
XX 06-OCT-2000 (first entry)
DT
XX
DE Human secreted protein 5' EST, SEQ ID NO: 23284.
XX
XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 9905-0122487.
XX
XX (GENSET) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
XX
XX
XX WPI; 2000-500381/45.
DR
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 23284; 71bp + CD-ROM; English.
PS
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or poly(A)+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
XX Sequence 186 BP; 54 A; 37 C; 38 G; 57 T; 0 other;
SQ
```

```
QY 120 aaagacattactctgtgtgagtaactccctaagtactcctcgtttttgtaagaacaagata 179
Db 61 aaagacattactctgtgtgagtaactccctaagtactcctcgtttttgtaagaacaagata 120
QY 180 cactcatgttaagagatgaagtcagatgaagtaactgttcgggtatcatatgaacaaga 239
Db 121 cactcatgttaagagatgaagtcagatgaagtaactgttcgggtatcatatgaacaaga 180
QY 240 gcc 242
Db 181 gcc 183

RESULT 4
AAS00048
ID AAS00048 standard; DNA; 31 BP.
XX
XX AAS00048;
AC
XX 14-MAY-2001 (first entry)
DT
XX
DE Human C2GnT3 PCR primer TSHC100.
XX
XX Human; C2GnT3; Thymus-related disorder; cancer; tumour; adenoma;
KM UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
KM malignant melanoma; breast cancer; cervical cancer; hypoadalivty;
KM hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
KM leukaemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;
KM acquired immunodeficiency syndrome; sepsis; wound healing; infection;
XX TSHC100; PCR primer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200114535-A2.
XX
XX 01-MAR-2001.
XX
XX 24-AUG-2000; 2000WO-DK00469.
XX
XX 24-AUG-1999; 9905-0150488.
XX
XX (SCHW/) SCHWIENTER T.
XX
XX (CLAU/) CLAUSEN H.
XX
XX Schwiientek T, Clausen H;
PI WPI; 2001-226615/23.
XX
XX
XX Example 2; Page 56; 97pp; English.
PS
XX
XX The sequence is PCR primer TSHC100 used to isolate DNA encoding
CC amino acids 39-453 of Human UDP-N-acetyl-galactosamine:Galactose-beta1,
CC 3-N-acetylgalactosamine-alpha-R beta1-6-N-acetylglucosaminyltransferase
CC (UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
CC C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
CC preparation of compositions for treating a conditions mediated by C2GnT3,
CC particularly a thymus-related disorder. C2GnT3, nucleic acids
CC encoding it and antibodies against it may also be used for in vitro
CC purposes related to scientific research, DNA synthesis and manufacture of
CC vectors, in the prognostic and diagnostic evaluation of conditions
CC associated with altered expression or activity of C2GnT3 or conditions
CC requiring modulation of C2GnT3, as well as in monitoring conditions by
CC detecting and localising the DNA and protein. Disorders such as tumours
CC (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
CC breast or cervix), hypoadalivty, hyperactivity, atrophy, enlargement of
CC thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
CC acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
CC sepsis, wound healing, acute and chronic infection, cell-mediated or
```

```

CC humour immunity, or TH1/TH2 imbalance, may be treated using these protein
CC or nucleic acid. The antibodies may be used to screen potential
CC therapeutic compounds to determine their effects on a conditions such as
CC thymus-related disorder or cancer, to determine the level of C2GnT3
CC expression in cells genetically engineered to produce C2GnT3, or to
CC detect and quantify polypeptides in a sample to determine their role in a
CC particular cellular events or pathological states and to diagnose and
CC treat such pathological states.
XX
SQ Sequence 31 BP; 10 A; 6 C; 7 G; 8 T; 0 other;

Query Match
Best Local Similarity 1.8%; Score 25; DB 22; Length 31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 114 tcgcgcaaaaagacattacttggtt 138
Db 7 tcgcgcaaaaagacattacttggtt 31

RESULT 5
AAS00046/c
ID AAS00046 standard; DNA; 21 BP.
AC AAS00046;
XX
DT 14-MAY-2001 (first entry)
DE Human C2GnT3 PCR primer TSHC96.
XX
KM Human: C2GnT3; Thymus-related disorder; cancer; tumour; adenoma;
KM UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
KM malignant melanoma; breast cancer; cervical cancer; hypocoactivity;
KM hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
KM leukaemia; lymphoma; immunosuppression; AIDS; Miskott-Aldrich syndrome;
KM acquired immunodeficiency syndrome; sepsis; wound healing; infection;
KM TSHC96; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200114535-A2.
XX
PD 01-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-DK00469.
XX
PR 24-AUG-1999; 99US-0150488.
XX
PA (SCHW/) SCHWIENTER T.
XX
PA (CLAU/) CLAUSEN H.
XX
PI Schlientek T, Clausen H;
XX
PI WPI; 2001-226615/23.
XX
DR New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
PT thymus-related disorders, cancers, tumours, immunosuppression.
XX
PS Example 1; Page 55; 97pp; English.
XX
XX The sequence represents PCR primer TSHC96 used to isolate genomic
CC DNA encoding Human UDP-N-acetyl-glucosamine:Galactose-beta1,
CC 3-N-acetylglactosamine-alpha-R beta1-6 N-acetylglucosaminyltransferase
CC (UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
CC C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
CC preparation of compositions for treating a conditions mediated by C2GnT3,
CC particularly a thymus-related disorder. C2GnT3, nucleic acids
CC encoding it and antibodies against it may also be used for in vitro
CC purposes related to scientific research, DNA synthesis and manufacture of
CC vectors, in the prognostic and diagnostic evaluation of conditions
CC associated with altered expression or activity of C2GnT3 or conditions

```

```

CC requiring modulation of C2GnT3, as well as in monitoring conditions by
CC detecting and localising the DNA and protein. Disorders such as tumours
CC (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
CC breast or cervix), hypocoactivity, hyperactivity, atrophy, enlargement of
CC thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
CC acquired immunodeficiency syndrome (AIDS), Miskott-Aldrich syndrome,
CC sepsis, wound healing, acute and chronic infection, cell-mediated or
CC humour immunity, or TH1/TH2 imbalance, may be treated using these protein
CC or nucleic acid. The antibodies may be used to screen potential
CC therapeutic compounds to determine their effects on a conditions such as
CC thymus-related disorder or cancer, to determine the level of C2GnT3
CC expression in cells genetically engineered to produce C2GnT3, or to
CC detect and quantify polypeptides in a sample to determine their role in a
CC particular cellular events or pathological states and to diagnose and
CC treat such pathological states.
XX
SQ Sequence 21 BP; 5 A; 7 C; 3 G; 6 T; 0 other;

Query Match
Best Local Similarity 1.5%; Score 21; DB 22; Length 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 750 tatgttgagacggtggaacc 770
Db 21 tatgttgagacggtggaacc 1

RESULT 6
AAS00047
ID AAS00047 standard; DNA; 32 BP.
AC AAS00047;
XX
DT 14-MAY-2001 (first entry)
DE Human C2GnT3 sequencing primer TSHC99.
XX
KM Human: C2GnT3; Thymus-related disorder; cancer; tumour; adenoma;
KM UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
KM malignant melanoma; breast cancer; cervical cancer; hypocoactivity;
KM hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
KM leukaemia; lymphoma; immunosuppression; AIDS; Miskott-Aldrich syndrome;
KM acquired immunodeficiency syndrome; sepsis; wound healing; infection;
KM TSHC99; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200114535-A2.
XX
PD 01-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-DK00469.
XX
PR 24-AUG-1999; 99US-0150488.
XX
PA (SCHW/) SCHWIENTER T.
XX
PA (CLAU/) CLAUSEN H.
XX
PI Schlientek T, Clausen H;
XX
PI WPI; 2001-226615/23.
XX
DR New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
PT thymus-related disorders, cancers, tumours, immunosuppression.
XX
PS Example 1; Page 55; 97pp; English.
XX
XX The sequence represents sequencing primer TSHC99 used to sequence a
CC cDNA encoding Human UDP-N-acetyl-glucosamine:Galactose-beta1,
CC 3-N-acetylglactosamine-alpha-R beta1-6 N-acetylglucosaminyltransferase
CC (UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or

```


Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 87 aaagcttcaatgcgaga 105
|||||
Db 163 aaagcttcaatgcgaga 181

RESULT 9
AAC41094
ID AAC41094 standard; DNA; 1127 bp.
XX
AC AAC41094;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 30623.
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP103405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128274.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130444.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 03-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147302.
PR 04-AUG-1999; 99US-0147304.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

```

PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149930.
PR 23-AUG-1999: 99US-0149930.
PR 23-AUG-1999: 99US-0149930.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 30-AUG-1999: 99US-0151080.
PR 31-AUG-1999: 99US-0151303.
PR 01-SEP-1999: 99US-0151438.
PR 07-SEP-1999: 99US-0151930.
PR 10-SEP-1999: 99US-0152363.
PR 13-SEP-1999: 99US-0153070.
PR 15-SEP-1999: 99US-0153758.
PR 16-SEP-1999: 99US-0154018.
PR 20-SEP-1999: 99US-0154039.
PR 22-SEP-1999: 99US-0154779.
PR 23-SEP-1999: 99US-0155139.
PR 24-SEP-1999: 99US-0155486.
PR 28-SEP-1999: 99US-0155659.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158829.
PR 08-OCT-1999: 99US-0158832.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159337.
PR 14-OCT-1999: 99US-0159637.
PR 18-OCT-1999: 99US-0159638.
PR 21-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161922.
PR 28-OCT-1999: 99US-0161993.
PR 28-OCT-1999: 99US-0162142.
PR 29-OCT-1999: 99US-0162142.

```

Query Match 1.4%; Score 19; DB 21; Length 1127;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 gagatgatgatgtgtgtg 313
 Db 94 gagatgatgatgtgtgtg 112

```

RESULT 10
AAS00049
ID AAS00049 standard; DNA; 20 BP.
XX
AC AAS00049:
XX
DT 14-MAY-2001 (first entry)
DE Human C2GnT3 PCR primer TSHC101.
XX
KW Human; C2GnT3; Thymus-related disorder; cancer; tumour; adenoma;
KW UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
KW malignant melanoma; breast cancer; cervical cancer; hypoadrenia;
KW hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
KW leukaemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;
KW acquired immunodeficiency syndrome; sepsis; wound healing; infection;
KW TSHC101; PCR primer; ss.
XX
OS Homo sapiens.
XX
MO 200114535-A2.
XX
PD 01-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-DK00469.
XX
PR 24-AUG-1999; 99US-0150488.
XX
PA (SCHW/) SCHWIENSTER T.
XX (CLAU/) CLAUSEN H.
XX
PI Schwiester T, Clausen H;
XX
DR WPI: 2001-226615/23.
XX
PR New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
PR useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
PR thymus-related disorders, cancers, tumours, immunosuppression.
XX
PS Example 1; Page 55; 97pp; English.
XX
CC The sequence represents PCR primer TSHC101 used to isolate genomic
CC DNA encoding Human UDP-N-acetyl-glucosamine:Galactose-beta1,
CC 3-N-acetylglucosamine-alpha-R beta1-6 N-acetylglucosaminyltransferase
CC (UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
CC C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
CC preparation of compositions for treating a conditions mediated by C2GnT3,
CC particularly a thymus-related disorder. C2GnT3, nucleic acids
CC encoding it and antibodies against it may also be used for in vitro
CC purposes related to scientific research, DNA synthesis and manufacture of
CC vectors, in the prognostic and diagnostic evaluation of conditions
CC associated with altered expression or activity of C2GnT3 or conditions
CC requiring modulation of C2GnT3, as well as in monitoring conditions by
CC detecting and localising the DNA and protein. Disorders such as tumours
CC (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
CC breast or cervix), hypoadrenia, hypoadrenia, atrophy, enlargement of
CC thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
CC acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
CC sepsis, wound healing, acute and chronic infection, cell-mediated or
CC humoral immunity, or TH1/TH2 imbalance, may be treated using these protein
CC or nucleic acid. The antibodies may be used to screen potential
CC therapeutic compounds to determine their effects on a conditions such as
CC thymus-related disorder or cancer, to determine the level of C2GnT3
CC expression in cells genetically engineered to produce C2GnT3, or to
CC detect and quantify polypeptides in a sample to determine their role in a
CC particular cellular events or pathological states and to diagnose and
CC treat such pathological states.
XX
SQ Sequence 20 BP; 5 A; 5 C; 4 G; 6 T; 0 other;

```

Query Match 1.3%; Score 18; DB 22; Length 20;

Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 495 tcgtgaagcactgtatc 512
|||||
Db 1 tcgtgaagcactgtatc 18

RESULT 11
AAH81918
ID AAH81918 standard; DNA; 255 BP.
AC AAH81918;
XX
XX
XX 21-SEP-2001 (first entry)
DE Rat differential transcription-associated CDNA SEQ ID 427.
XX
XX
XX Differential transcription; human; rat; tumour cell; cytostatic;
KM Ras modulator; Class II tumour suppressor gene; gene therapy; ss.
XX
XX
XX Rattus sp.
OS
XX
XX WO200157058-A2.
XX
XX
XX 09-AUG-2001.
XX
XX
XX 31-JAN-2001; 2001WO-EP01003.
XX
XX 31-JAN-2000; 2000DE-1004102.
XX
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;
PI Grips M, Hellriegel M, Schmitz A, Sers C;
XX
XX WPI: 2001-483415/52.
XX
XX
XX Nucleic acids differentially expressed between tumor and normal cells,
PT useful for diagnosis or therapy of tumors and for screening active
PT agents -
XX
XX
XX Claim 6; Page 495; 579pp; German.
XX
XX
XX This invention describes a nucleic acid (I) with differential expression
CC between tumour and normal cells and which has cytostatic activity. (I)
CC work as modulators of Ras activity by inducing expression of tumour
CC suppressor genes. (I), and polypeptides encoded by them, are useful as
CC targets for diagnosis or therapy and in screening to determine the
CC effects of an active compound (potential pharmaceutical) on a cell line,
CC particularly for diagnosis and treatment of tumors, especially by
CC modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
CC methods) or by modulating the amount and/or location of (I)-encoded
CC polypeptides (by administration of the polypeptide or its activator,
CC antibody (optionally as a conjugate) or inhibitor). The method allows
CC identification of many Class II tumour suppressor genes (i.e. genes that
CC are not primary targets for tumour-initiating mutations).
CC AAH81492-AAH83376 represent the human and rat derived nucleic acid
CC fragments described in the method of the invention.
XX
XX
XX Sequence 255 BP; 89 A; 57 C; 44 G; 65 T; 0 other;
SQ

Query Match 1.3%; Score 18; DB 22; Length 255;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1276 gaaaagcttgaagacag 1293
|||||
Db 237 gaaaagcttgaagacag 254

RESULT 12

AAZ13168/c
ID AAZ13168 standard; CDNA; 300 BP.
XX
XX
XX AAZ13168;
XX
XX
XX 12-OCT-1999 (first entry)
DE
XX
XX
XX Human gene expression product CDNA sequence SEQ ID NO:637.
XX
XX
XX Human; gene: gene expression product; diagnosis; therapy; probe;
KM detection; mapping; tissue typing; profiling; forensic; cancer;
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
XX
XX Homo sapiens.
OS
XX
XX WO9938972-A2.
XX
XX
XX 05-AUG-1999.
XX
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX
XX
XX 03-APR-1998; 98US-0080666.
XX
XX 28-JAN-1998; 98US-0072910.
XX
XX 24-FEB-1998; 98US-0075954.
XX
XX 31-MAR-1998; 98US-0080114.
XX
XX 03-APR-1998; 98US-0080515.
XX
XX
XX (CHIR) CHIRON CORP.
XX
XX (HYSE-) HYSEQ INC.
XX
XX
XX Cirkenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Imis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat T;
PI Lamsom G, Lesnikowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Grain B, Sudduth-Klinger J, Williams LT;
XX
XX
XX WPI: 1999-494092/41.
XX
XX
XX Novel human genes and their expression products which are
PT differentially expressed in different cell types
PT
XX
XX
XX Claim 1; Page 807-808; 2479pp; English.

The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
XX
XX Sequence 300 BP; 71 A; 87 C; 62 G; 68 T; 12 other;
SQ

Query Match 1.3%; Score 18; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1003 gatgagcactttggct 1020

Db 269 GATGAGACTTTGGCT 252

RESULT 13
AAFI1803/c
ID AAFI1803 standard; cDNA; 528 BP.

XX AAFI1803:

AC 13-MAR-2001 (first entry)

DE Aspergillus niger EST SEQ ID NO:4326.

XX Multiple gene expression; filamentous fungal cell; EST;
KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KM culture condition; environmental stress; spore morphogenesis;
KM metabolic pathway engineering; catabolic pathway engineering; ss.

OS Aspergillus niger.

PN WO200056762-A2.

XX 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.

PR 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;

DR WPI: 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags.

PS Claim 87; Page 1862; 3161pp; English.

XX The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stresses, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAFI1803 to AAFI1803 represents ESTs from
CC Fusarium venenatum; AAFI1803 to AAFI1803 represents ESTs from
CC niger; AAFI1803 to AAFI1803 represents ESTs from Aspergillus
CC niger; AAFI1803 to AAFI1803 represents ESTs from Trichoderma reesei; and
CC AAFI1803 to AAFI1803 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

XX Sequence 528 BP; 129 A; 129 C; 135 G; 135 T; 0 other;

Query Match 1.3%; Score 18; DB 21; Length 528;
Best Local Similarity 100.0%; Pred. No. 82;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 465 CCAGCACAATATTACTG 482
Db 326 CCAGCACAATATTACTG 309

RESULT 14
AAI12009
ID AAI12009 standard; DNA; 592 BP.

XX AAI12009;

DE 08-OCT-1999 (first entry)

XX Neisseria meningitidis partial ORF1 sequence.

DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KM treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.

OS Neisseria meningitidis.

PN WO924578-A2.

XX 20-MAY-1999.

PF 09-OCT-1998; 98WO-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0022516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

XX (CHIR-) CHIRON SPA.

PA Grandi G, Masiagnani V, Pizzi M, Rappuoli R, Scarlato V;

PI WPI: 1999-327407/27.

DR P-PSDB: AAY38541.

XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for

PT diagnosis, treatment and prevention of infection

PS Claim 9; Page 101; 524pp; English.

CC Nucleotide sequences AAI1972-212358 represent open reading frames

CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode

CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their

CC fragments, their nucleic acids and antibodies are used for diagnosis,

CC prevention (as vaccines) or treatment of Neisseria infections,

CC such as meningitis, septicemia and gonorrhea. Both organisms

CC are closely related. Fragments of the nucleic acids are useful

CC as hybridisation probes and antisense reagents.

XX Sequence 592 BP; 124 A; 190 C; 130 G; 145 T; 3 other;

Query Match 1.3%; Score 18; DB 20; Length 592;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 ttatctctgttttaacc 66
Db 128 ttatctctgttttaacc 145

RESULT 15
AAI1331
ID AAI1331 standard; DNA; 592 BP.

AC AA81331;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis Menb polynucleotide sequence ORF number 10.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; Menb; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200022430-A2.
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-US23573.
 XX
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI -Prazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M.
 XX
 DR WPI: 2000-318079/27.
 XX
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX
 PS Disclosure: Page 202; 1760pp; English.
 XX
 XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AA81453 to AA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences: AA81260 to AA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AA81254 to
 CC AA81259 and AA81304 to AA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AA81322 to
 CC AA81452 represent Neisseria meningitidis Menb polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neissariae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 SQ Sequence 592 BP; 124 A; 190 C; 130 G; 145 T; 3 other;

Query Match 1.3%; Score 18; DB 21; Length 592;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 ttcatcctgttttaacc 66
 ||||||||||||||||
 DB 128 ttcatcctgttttaacc 145

RESULT 16
 AA212010
 ID AA212010 standard; DNA; 807 BP.
 XX
 AC AA212010;
 XX
 DT 08-OCT-1999 (first entry)
 XX
 DE Complete ORF17 sequence of Neisseria meningitidis.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9924578-A2.
 PD 20-MAY-1999.
 XX
 PF 09-OCT-1998; 98WO-IB01665.
 XX
 PR 01-SEP-1998; 98GB-0019016.
 PR 06-NOV-1997; 97GB-0023516.
 PR 14-NOV-1997; 97GB-0024150.
 PR 18-NOV-1997; 97GB-0024386.
 PR 27-NOV-1997; 97GB-0025158.
 PR 10-DEC-1997; 97GB-0026147.
 PR 14-JAN-1998; 98GB-0000759.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
 PI P-PSDB; AAY38542.
 DR WPI: 1999-327407/27.
 XX
 PS Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
 PT diagnosis, treatment and prevention of infection
 PT
 XX Claim 9; Page 102; 524pp; English.
 XX
 CC Nucleotide sequences AA211972-212358 represent open reading frames
 CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
 CC fragments, their nucleic acids and antibodies are used for diagnosis,
 CC prevention (as vaccines) or treatment of Neisseria infections,
 CC such as meningitis, septicemia and gonorrhea. Both organisms
 CC are closely related. Fragments of the nucleic acids are useful
 CC as hybridisation probes and antisense reagents.
 XX
 SQ Sequence 807 BP; 153 A; 252 C; 192 G; 209 T; 1 other;

Query Match 1.3%; Score 18; DB 20; Length 807;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 ttcatcctgttttaacc 66
 ||||||||||||||||
 DB 343 ttcatcctgttttaacc 360

RESULT 17
 AA212011
 ID AA212011 standard; DNA; 807 BP.
 XX
 AC AA212011;
 XX
 DT 08-OCT-1999 (first entry)
 XX
 DE Complete ORF17 sequence of Neisseria meningitidis strain A.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

DR P-PSDB; AAW34332.
 XX Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection
 PT e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
 PT golden mosaic geminivirus
 XX
 PS Example 3.3; Page 57-58; 132pp; English.
 XX
 CC This genomic DNA sequence includes the open reading frame of the
 CC wild-type AC1 gene of tomato mottle virus (TOMOV), a geminivirus
 CC that has a bipartite genome. The AC1 gene must be expressed for
 CC efficient replication of the two genomic components, DNA-A and
 CC DNA-B. It encodes a protein (see AAW34336) having a DNA binding
 CC site specific to the DNA-A common region, a DNA nicking activity,
 CC and an NTP binding activity. The invention involves production of
 CC transgenic plants containing DNA comprising AC1 or CI wild-type or
 CC mutant sequences that negatively interfere in trans with
 CC geminiviral replication during infection. Such transgenic plants
 CC are resistant to viral infection. The AC1/CI genes are especially
 CC from TOMOV, tomato yellow leaf curl virus or bean golden mosaic
 CC geminivirus (see AAT93282-93) and encode polypeptides (see AAW34324-35)
 CC that have mutations in the highly conserved DNA-nicking domain
 CC and/or the NTP-binding domain.
 CC
 SO Sequence 1160 BP; 360 A; 276 C; 257 G; 267 T; 0 other;
 Query Match 1.3%; Score 18; DB 18; Length 1160;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1047 tggggagattccagatc 1064
 Db 380 tggggagattccagatc 397
 RESULT 20
 AAT93285
 ID AAT93285 standard; DNA; 1166 BP.
 AC AAT93285;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Tomato mottle virus AC1 mutant gene.
 XX
 KW Geminivirus; TOMOV-AC1d1m23; AC1 gene; transdominant mutation;
 KM transgenic plant; disease resistance; ss; cyclic; circular.
 XX
 OS Tomato mottle virus isolate Florida.
 OS Synthetic.
 XX
 FT Key Location/Qualifiers
 FT CDS 44..439
 FT /*tag= a
 FT
 XX WO9739110-A1.
 PN
 XX 23-OCT-1997.
 PD
 XX 15-APR-1997; 97WO-US06300.
 PF
 XX 16-APR-1996; 96US-0015517.
 PR
 XX (SEMI-) SEMINIS VEGETABLE SEEDS INC.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT;
 XX WPI; 1997-526447/48.
 DR P-PSDB; AAW34332.
 XX

PT Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection
 PT e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
 PT golden mosaic geminivirus
 XX
 PS Claim 11; Page 72-73; 132pp; English.
 XX
 CC This DNA sequence comprises a mutated AC1 gene of tomato mottle
 CC virus (TOMOV). It carries an inserted 4-base Sau3a site that
 CC shifts the AC1 gene translation reading frame resulting in the
 CC expression of a truncated AC1 protein (see AAW34327). The AC1 gene
 CC (see also AAT93294) must be expressed for efficient replication of
 CC the two genomic components, DNA-A and DNA-B, of the bipartite TOMOV
 CC genome. The invention involves production of transgenic plants
 CC containing DNA comprising geminivirus AC1 or CI wild-type or mutant
 CC sequences that negatively interfere in trans with geminiviral
 CC replication during infection. Such transgenic plants are resistant
 CC to viral infection. The AC1/CI genes are especially from TOMOV,
 CC tomato yellow leaf curl virus or bean golden mosaic geminivirus
 CC (see AAT93282-93) and encode polypeptides (see AAW34324-35) that have
 CC mutations in the highly conserved DNA-nicking and/or the NTP-binding
 CC domains.
 CC
 SO Sequence 1166 BP; 361 A; 277 C; 260 G; 268 T; 0 other;
 Query Match 1.3%; Score 18; DB 18; Length 1166;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1047 tggggagattccagatc 1064
 Db 380 tggggagattccagatc 397
 RESULT 21
 AAT93282
 ID AAT93282 standard; DNA; 1169 BP.
 AC AAT93282;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Tomato mottle virus AC1 mutant TOMOV-AC1d1m gene.
 XX
 KW Geminivirus; TOMOV-AC1d1m; AC1 gene; transdominant mutation;
 KM transgenic plant; disease resistance; ss; cyclic; circular.
 XX
 OS Tomato mottle virus isolate Florida.
 OS Synthetic.
 XX
 FT Key Location/Qualifiers
 FT CDS 44..1129
 FT /*tag= a
 FT
 XX WO9739110-A1.
 PN
 XX 23-OCT-1997.
 PD
 XX 15-APR-1997; 97WO-US06300.
 PF
 XX 16-APR-1996; 96US-0015517.
 PR
 XX (SEMI-) SEMINIS VEGETABLE SEEDS INC.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT;
 XX WPI; 1997-526447/48.
 DR P-PSDB; AAW34324.
 XX
 PT Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection

PT e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
 PT golden mosaic geminivirus
 XX
 PS Claim 11: Page 60-62; 132pp; English.
 CC This DNA sequence comprises a transdominant lethal mutant,
 CC designated TOMOV-ACIdim, of the AC1 gene of tomato mottle virus
 CC virus (TOMOV). It encodes an AC1 protein (see AAW34325) that carries
 CC mutations in its NTP-binding domains. The AC1 gene (see also
 CC AAT93294) must be expressed for efficient replication of the two
 CC genomic components, DNA-A and DNA-B, of the bipartite TOMOV genome.
 CC The invention involves production of transgenic plants containing
 CC DNA comprising geminivirus AC1 or CI wild-type or mutant sequences
 CC that negatively interfere in trans with geminiviral replication
 CC during infection. Such transgenic plants are resistant to viral
 CC infection. The AC1/CI genes are especially from TOMOV, tomato
 CC yellow leaf curl virus or bean golden mosaic geminivirus (see
 CC AAT93282-93) and encode polypeptides (see AAW34324-35) that have
 CC mutations in the highly conserved DNA-nicking and/or the NTP-binding
 CC domains.
 XX
 SQ Sequence 1169 BP; 363 A; 281 C; 255 G; 270 T; 0 other;
 Query Match 1.3%; Score 18; DB 18; Length 1169;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1047 tggggagattccagatc 1064
 ||||||||||||||||
 Db 380 tggggagattccagatc 397
 RESULT 22
 AAT93283
 ID AAT93283 standard; DNA; 1169 BP.
 AC AAT93283;
 XX
 DT 27-APR-1998 (first entry)
 DE Tomato mottle virus AC1 mutant TOMOV-ACIdim1 gene.
 XX
 KW Geminivirus; TOMOV-ACIdim1; AC1 gene; transdominant mutation;
 KW transgenic plant; disease resistance; ss; cyclic; circular.
 XX
 OS Tomato mottle virus isolate Florida.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 44..1129
 FT /*tag= a
 PN MO9739110-A1.
 PD 23-OCT-1997.
 XX
 PF 15-APR-1997; 97WO-US06300.
 XX
 PR 16-APR-1996; 96US-0015517.
 XX
 PA (SEMI-) SEMINIS VEGETABLE SEEDS INC.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT;
 XX
 DR WPI: 1997-526447/48.
 DR P-PSDB: AAW34325.
 XX
 PT Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection
 PT e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
 PT golden mosaic geminivirus

XX
 PS Claim 11: Page 64-65; 132pp; English.
 CC This DNA sequence comprises a transdominant lethal mutant,
 CC designated TOMOV-ACIdim, of the AC1 gene of tomato mottle virus
 CC virus (TOMOV). It encodes an AC1 protein (see AAW34325) that carries
 CC a mutation in its NTP-binding domains. The AC1 gene (see also
 CC AAT93294) must be expressed for efficient replication of the two
 CC genomic components, DNA-A and DNA-B, of the bipartite TOMOV genome.
 CC The invention involves production of transgenic plants containing
 CC DNA comprising geminivirus AC1 or CI wild-type or mutant sequences
 CC that negatively interfere in trans with geminiviral replication
 CC during infection. Such transgenic plants are resistant to viral
 CC infection. The AC1/CI genes are especially from TOMOV, tomato
 CC yellow leaf curl virus or bean golden mosaic geminivirus (see
 CC AAT93282-93) and encode polypeptides (see AAW34324-35) that have
 CC mutations in the highly conserved DNA-nicking and/or the NTP-binding
 CC domains.
 XX
 SQ Sequence 1169 BP; 361 A; 280 C; 258 G; 270 T; 0 other;
 Query Match 1.3%; Score 18; DB 18; Length 1169;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1047 tggggagattccagatc 1064
 ||||||||||||||||
 Db 380 tggggagattccagatc 397
 RESULT 23
 AAT93284
 ID AAT93284 standard; DNA; 1169 BP.
 AC AAT93284;
 XX
 DT 27-APR-1998 (first entry)
 DE Tomato mottle virus AC1 mutant TOMOV-ACIdim23 gene.
 XX
 KW Geminivirus; TOMOV-ACIdim23; AC1 gene; transdominant mutation;
 KW transgenic plant; disease resistance; ss; cyclic; circular.
 XX
 OS Tomato mottle virus isolate Florida.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 44..1129
 FT /*tag= a
 PN MO9739110-A1.
 PD 23-OCT-1997.
 XX
 PF 15-APR-1997; 97WO-US06300.
 XX
 PR 16-APR-1996; 96US-0015517.
 XX
 PA (SEMI-) SEMINIS VEGETABLE SEEDS INC.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT;
 XX
 DR WPI: 1997-526447/48.
 DR P-PSDB: AAW34326.
 XX
 PT Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection
 PT e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
 PT golden mosaic geminivirus
 PS Claim 11: Page 67-69; 132pp; English.

XX This DNA sequence comprises a transdominant lethal mutant.
CC designated TOMOV-NC1d1m23, of the AC1 gene of tomato mottle virus
CC virus (TOMOV). It encodes an AC1 protein (see AAW4326) that carries
CC 2 mutations in an NTP-binding domain. The AC1 gene (see also
CC AAT93294) must be expressed for efficient replication of the two
CC genomic components, DNA-A and DNA-B, of the bipartite TOMOV genome.
CC The invention involves production of transgenic plants containing
CC DNA comprising geminivirus AC1 or CI wild-type or mutant sequences
CC that negatively interfere in trans with geminiviral replication
CC during infection. Such transgenic plants are resistant to viral
CC infection. The AC1/CI genes are especially from TOMOV, tomato
CC yellow leaf curl virus or bean golden mosaic geminivirus (see
CC AAT93282-93) and encode polypeptides (see AAW34324-35) that have
CC mutations in the highly conserved DNA-nicking and/or the NTP-binding
CC domains.
XX
SQ Sequence 1169 BP; 364 A; 278 C; 257 G; 270 T; 0 other;

Query Match 1.3%; Score 18; DB 18; Length 1169;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 tgggagattccagatc 1064
Db 380 tgggagattccagatc 397
|||||

RESULT 24
AAC42240/C
ID AAC42240 standard; DNA; 1463 BP.
XX
AC AAC42240;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 34807.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 200EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0127462.
PR 01-APR-1999; 99US-0128234.
PR 06-APR-1999; 99US-0128714.
PR 08-APR-1999; 99US-0129645.
PR 16-APR-1999; 99US-0130077.
PR 19-APR-1999; 99US-0130449.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130691.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

KW Hybridisation assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135323.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136352.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138340.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

```

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150584.
PR 27-AUG-1999; 99US-0151063.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156358.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

```

```

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161930.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.3%; Score 18; DB 21; Length 1463;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 18; Conservatve 0; Mismatches 0; Indels 0; Gaps 0;

Qy 474 tattactgcatacatta 491
Db 750 TATTACTGCATCATTA 733

RESULT 25
ID AAA81548 standard; DNA; 2456 BP.
AC AAA81548;
XX 04-DEC-2000 (first entry)
DT 04-DEC-2000 (first entry)
XX N. meningitidis partial DNA sequence gnm_95 SEQ ID NO:95.
DE N. meningitidis partial DNA sequence gnm_95 SEQ ID NO:95.
XX N. meningitidis partial DNA sequence gnm_95 SEQ ID NO:95.
KM Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
KM Meningococcus B; Menb; ds.
XX Neisseria meningitidis.
OS Neisseria meningitidis.
PN WO200022430-A2.
XX 20-APR-2000.
PD 20-APR-2000.
XX 08-OCT-1999; 99WO-US23573.
PF 08-OCT-1999; 99WO-US23573.
XX 09-OCT-1999; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX (CHIR) CHIRON CORP.
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Malignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;
PI Rappuoli R, Pianza M;
XX WPI: 2000-318079/27.
DR 2000-318079/27.
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX Claim 7; Page 1487-1488; 1760pp; English.
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA8153 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis Menb polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a

```

CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B, against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 CC
 XX Sequence 2456 BP; 553 A; 753 C; 565 G; 585 T; 0 other;
 SQ
 Query Match 1.3%; Score 18; DB 21; Length 2456;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 49 ttcacccgttttttaacc 66
 DB 1805 ttcacccgttttttaacc 1822
 RESULT 26
 AA193309/C
 ID AA193309 standard; DNA; 2602 BP.
 XX
 AC AA193309;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Tomato mottle virus full-length A-component clone.
 XX
 KW Geminivirus; TOMOV; AC1 gene; transdominant mutation;
 KW Transgenic plant; disease resistance; ss: cyclic; circular.
 XX
 OS Tomato mottle virus isolate Florida.
 XX
 PN WO9739110-A1.
 XX
 PD 23-OCT-1997.
 XX
 PF 15-APR-1997; 97WO-US06300.
 XX
 PR 16-APR-1996; 960S-0015517.
 XX
 PA (SEMI-) SEMINIS VEGETABLE SEEDS INC.
 XX (MISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT.
 XX
 DR WPI; 1997-526447/48.
 XX
 PT Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection
 PT e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
 PT golden mosaic geminivirus
 XX
 PS Example 3.1; Page 76-77; 132pp; English.
 XX
 CC This genomic DNA sequence comprises a full-length A-component clone
 CC of tomato mottle virus (TOMOV), a geminivirus that has a bipartite
 CC genome comprising DNA-A and DNA-B. It was isolated from TOMOV
 CC infected Nicotiana benthamiana and tomato plant DNA by restriction
 CC digestion. The DNA-B component (see AA193310) was also isolated.
 CC TOMOV DNA-A contains the AC1 gene (see AA193294) that must be
 CC expressed for efficient replication of DNA-A and DNA-B. The
 CC invention involves production of transgenic plants containing DNA
 CC comprising AC1 or CI wild-type or mutant sequences that negatively
 CC interfere in trans with geminiviral replication during infection.

CC Such transgenic plants are resistant to viral infection.
 XX
 SQ Sequence 2602 BP; 671 A; 561 C; 586 G; 784 T; 0 other;
 Query Match 1.3%; Score 18; DB 18; Length 2602;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1047 tggggagatttcagacc 1064
 DB 2273 TGGGGAGATTTCAGACC 2256
 RESULT 27
 AA257860/C
 ID AA257860 standard; cDNA; 2843 BP.
 XX
 AC AA257860;
 XX
 DT 11-APR-2000 (first entry)
 XX
 DE Protein regulating gene expression PRGE-22 cDNA clone 2998209.
 XX
 KW Protein regulating gene expression; PRGE-22; human;
 KW cancer; inflammation; anticancer; antitumor; antiinflammatory;
 KW zinc finger; apoptosis inhibitor; antiapoptotic; gene therapy;
 KW diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 377..1822
 FT /*tag= a
 XX
 PN WO9964596-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13281.
 XX
 PR 12-JUN-1998; 98US-0089029.
 PR 29-JUL-1998; 98US-0094575.
 PR 14-OCT-1998; 98US-0104624.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Lal P, Yue H, Tang YT, Hillman JL, Bandman O, Corley NC;
 PI Guegler KJ, Gorgone GA, Baughn MR, Patterson C, Lu DM;
 XX
 DR WPI; 2000-116543/10.
 XX P-PSDB; AA158629.
 XX
 PT New human polypeptides that regulate gene expression, for treatment,
 PT prevention and diagnosis of, e.g. cancer.
 XX
 PS Claim 9; Page 138; 150pp; English.
 XX
 CC The present sequence is that of Incyte clone 2998209 encoding new
 CC human protein regulating gene expression PRGE-22 (see AA158629). The
 CC cDNA was initially isolated from ovarian tumor tissue cDNA library
 CC OVAR107, and the full-length sequence assembled from overlapping
 CC clones from a number of libraries. PRGE-22 is expressed in
 CC reproductive, nervous and gastrointestinal tissues associated with
 CC cancer, inflammation and foetal diseases, disorders or conditions.
 CC It is characterized as a zinc finger protein and inhibitor of
 CC apoptosis. The invention provides PRGE polypeptides (see AA158608-38)
 CC and polynucleotides (see AA257839-69), expression vectors, host cells,
 CC antibodies, agonists and antagonists. It also provides methods for
 CC diagnosing, treating or preventing disorders associated with
 CC expression of PRGE. Polynucleotides are also used as sources of
 CC probes and primers for diagnosis and monitoring of disease, also
 CC for detecting related sequences and in gene mapping.

XX Sequence 2843 BP; 835 A; 537 C; 662 G; 809 T; 0 other;
SQ

Query Match 1.3%; Score 18; DB 21; Length 2843;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ttctcatcctgtttttaa 64
|||||
Db 476 TTTTCATCCGTGTTTAA 459

RESULT 28
AAFA4664
ID AAF44664 standard; cDNA: 3864 BP.
XX
XX AAF44664;
AC
DT 27-MAR-2001 (first entry)
XX
XX Novel protein kinase cDNA, SEQ ID NO: 44.
DE
XX Human: mouse; protein kinase; antiarthritic; atherosclerotic; osteopathic;
KM immunosuppressive; cardiant; antiinflammatory; antiasthmatic;
KM dermatological; antidiabetic; antifertility; gene therapy; vaccine;
KM immune disorder; cardiovascular disease; neurodegenerative disease;
KM cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KM inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
XX
XX Homo sapiens.
OS
PN WO200073469-A2.
XX
XX 07-DEC-2000.
PD
XX 26-MAY-2000; 2000MO-US14842.
PF
XX 28-MAY-1999; 9905-0136503.
PR
XX (SUGEN-) SUGEN INC.
PA
XX Plowman GD, Martinez R, Whyte D, Sudersanam S;
PI WPI, 2001-032161/04.
DR P-PSDB: AAB65637.
XX
XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers -
XX
XX Example 1; Fig 2; 310pp; English.
PS
XX The present sequence encodes a novel protein kinase. The nucleic acids
CC and the protein kinases they encode may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative-stress related disorders, chronic inflammatory bowel disease,
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC reproductive disorders.
XX
SQ Sequence 3864 BP; 1048 A; 995 C; 995 G; 826 T; 0 other;

Query Match 1.3%; Score 18; DB 22; Length 3864;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 tcatgacttgaggatg 301
|||||
Db 3368 tcatgacttgaggatg 3385

RESULT 29
AAC75062/C
ID AAC75062 standard; cDNA: 4982 BP.
XX
XX AAC75062;
AC
DT 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF617 polynucleotide sequence SEQ ID NO:1233.
DE
XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
KM vulnerable; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KM hypotensive; dermatological; immunosuppressive; antiinflammatory;
KM antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
KM antiaesthetic; gene therapy; cancer; proliferative disorder; hypertension;
KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KM cholesterol ester storage; systemic lupus erythematosus; infection;
KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
KM thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
OS
XX WO200058473-A2.
PN
XX 05-OCT-2000.
PD
XX 31-MAR-2000; 2000MO-US08621.
PF
XX 31-MAR-1999; 9905-0127607.
PR 02-APR-1999; 9905-0127636.
PR 05-APR-1999; 9905-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Shinkets RA, Leach M;
PI WPI, 2000-602362/57.
DR P-PSDB: AAB40853.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 1091-1094; 5507pp; English.
PS
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antihypoid; and antianesthetic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 4982 BP; 1475 A; 901 C; 1096 G; 1508 T; 2 other;

Query Match 1.3%; Score 18; DB 21; Length 4982;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ttttcacccgttttta 64
DB 1189 ttttcacccgttttta 1172

RESULT 30
ID AAX89220 standard; DNA: 5355 BP.
XX
AC AAX89220:
XX
DT 15-SEP-1999 (first entry)
XX
DE Seq ID No: 28 of WO9933961.
XX
KM DRAK1: DAP kinase related apoptosis inducing kinase; human;
KM apoptosis; breast cancer; ovarian cancer; lymphoma; autoimmune disease;
KM viral infection; adenovirus; poxvirus; HIV; Alzheimer's disease;
KM Parkinson's disease; arteriosclerosis; alcoholism; rheumatoid arthritis;
KM diabetes; ss.
XX
OS Homo sapiens.
XX
PN WO9933961-A1.
XX
PD 08-JUL-1999.
XX
PF 25-DEC-1998; 98WO-JP05974.
XX
PR 17-APR-1998; 98JP-0108150.
PR 26-DEC-1997; 97JP-0367640.
PR 26-DEC-1997; 97JP-0367641.
PR 17-APR-1998; 98JP-0108149.
XX
PA (ASAH) ASAH KASEI KOGYO KK.
XX
PI Akira S, Kawai T;
XX
DR WPI, 1999-430239/36.
DR P-PDB; AAT7163.
XX
PT New kinase with apoptosis induction activity useful in the treatment
PT of cancer, autoimmune diseases and viral infections
XX
PS Disclosure; Page 157-166; 180pp; Japanese.
XX
CC The invention provides kinases DRAK1 and DRAK2 (DAP kinase related
CC apoptosis inducing kinase) having apoptosis inducing activity. The
CC kinases can be expressed recombinantly by transforming host cells with
CC vectors comprising the nucleic acids encoding the kinases. The kinases
CC are useful in the treatment, prevention, diagnosis and investigation of
CC diseases with which apoptosis is associated, such as hormonally regulated
CC cancer (such as breast cancer, ovarian cancer, lymphoma); autoimmune
CC diseases; viral infections (such as herpes, adenovirus, poxvirus, HIV);
CC Alzheimer's disease; Parkinson's disease; arteriosclerosis; alcoholism;
CC rheumatoid arthritis; and diabetes.

XX
SQ Sequence 5355 BP; 1515 A; 1278 C; 1302 G; 1260 T; 0 other;

Query Match 1.3%; Score 18; DB 20; Length 5355;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 tcatgactggagagatg 301
DB 3491 tcatgactggagagatg 3508

RESULT 31
ID AAX89221/C
XX
AC AAX89221 standard; DNA: 5355 BP.
XX
DT 15-SEP-1999 (first entry)
XX
DE Seq ID No: 30 of WO9933961.
XX
KM DRAK1: DAP kinase related apoptosis inducing kinase; human;
KM apoptosis; breast cancer; ovarian cancer; lymphoma; autoimmune disease;
KM viral infection; adenovirus; poxvirus; HIV; Alzheimer's disease;
KM Parkinson's disease; arteriosclerosis; alcoholism; rheumatoid arthritis;
KM diabetes; ss.
XX
OS Homo sapiens.
XX
PN WO9933961-A1.
XX
PD 08-JUL-1999.
XX
PF 25-DEC-1998; 98WO-JP05974.
XX
PR 17-APR-1998; 98JP-0108150.
PR 26-DEC-1997; 97JP-0367640.
PR 26-DEC-1997; 97JP-0367641.
PR 17-APR-1998; 98JP-0108149.
XX
PA (ASAH) ASAH KASEI KOGYO KK.
XX
PI Akira S, Kawai T;
XX
DR WPI, 1999-430239/36.
XX
PT New kinase with apoptosis induction activity useful in the treatment
PT of cancer, autoimmune diseases and viral infections
XX
PS Disclosure; Page 173-176; 180pp; Japanese.
XX
CC The invention provides kinases DRAK1 and DRAK2 (DAP kinase related
CC apoptosis inducing kinase) having apoptosis inducing activity. The
CC kinases can be expressed recombinantly by transforming host cells with
CC vectors comprising the nucleic acids encoding the kinases. The kinases
CC are useful in the treatment, prevention, diagnosis and investigation of
CC diseases with which apoptosis is associated, such as hormonally regulated
CC cancer (such as breast cancer, ovarian cancer, lymphoma); autoimmune
CC diseases; viral infections (such as herpes, adenovirus, poxvirus, HIV);
CC Alzheimer's disease; Parkinson's disease; arteriosclerosis; alcoholism;
CC rheumatoid arthritis; and diabetes.

Query Match 1.3%; Score 18; DB 20; Length 5355;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 tcatgactggagagatg 301
|||||

Db 1865 TCATTGACTTGAGGATG 1848

RESULT 32

AAZ39404
ID AAZ39404 standard; cDNA to mRNA; 5355 BP.

AC AAZ39404;

DT 23-FEB-2000 (first entry)

XX Human Trad protein encoding nucleotide sequence.

XX Trad protein; phosphoenzymatic activity; diagnosis; Rho subfamily;
KW skeletal muscle; human; ss.

OS Homo sapiens.

PN JPL1290083-A.

PD 26-OCT-1999.

PF 17-APR-1998; 98JP-0108151.

PR 17-APR-1998; 98JP-0108151.

PA (ASAH) ASAH KASEI KOGYO KK.

DR WPI; 2000-016982/02.

DR P-PSDB; AAY56781.

PT A new protein Trad or its salts - useful for the screening of diseases
involving the Rho subfamily on the function of skeletal muscles

PS Claim 4; Page 13-18; 25pp; Japanese.

CC This DNA encodes a human protein Trad. The protein is useful for the
screening and the evaluation of an inducer or an inhibitor of

CC phosphoenzymatic activity of Trad and for the diagnosis of diseases
participated by Rho subfamily on the function of skeletal muscles.

CC Sequence 5355 BP; 1515 A; 1278 C; 1302 G; 1260 T; 0 other;

Query Match 1.3%; Score 18; DB 21; Length 5355;
Best Local Similarity 100.0%; Pred. No. 75;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 tcattgacttgaggatg 301
|||||

Db 3491 tcattgacttgaggatg 3508

RESULT 33

AAZ39405/C
ID AAZ39405 standard; cDNA to mRNA; 5355 BP.

AC AAZ39405;

DT 23-FEB-2000 (first entry)

XX Human Trad protein DNA complementary nucleotide sequence.

XX Trad protein; phosphoenzymatic activity; diagnosis; Rho subfamily;
KW skeletal muscle; human; complementary; ss.

OS Homo sapiens.

PN JPL1290083-A.

PD 26-OCT-1999.

PF 17-APR-1998; 98JP-0108151.

XX 17-APR-1998; 98JP-0108151.

XX (ASAH) ASAH KASEI KOGYO KK.

DR WPI; 2000-016982/02.

PT A new protein Trad or its salts - useful for the screening of diseases
involving the Rho subfamily on the function of skeletal muscles

PS Claim 9; Page 19-20; 25pp; Japanese.

CC The invention provides a human protein Trad. The protein is useful for
the screening and the evaluation of an inducer or an inhibitor of

CC phosphoenzymatic activity of Trad and for the diagnosis of diseases
participated by Rho subfamily on the function of skeletal muscles. The

CC present sequence represents the complementary sequence of the Trad DNA.

CC Sequence 5355 BP; 1260 A; 1302 C; 1278 G; 1515 T; 0 other;

Query Match 1.3%; Score 18; DB 21; Length 5355;
Best Local Similarity 100.0%; Pred. No. 75;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 tcattgacttgaggatg 301
|||||

Db 1865 TCATTGACTTGAGGATG 1848

RESULT 34

AAD02701
ID AAD02701 standard; DNA; 27150 BP.

AC AAD02701;

DT 02-MAY-2001 (first entry)

DE Human glycosyl sulfotransferase-6 (GST-6) genomic DNA #1.

KW Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive;
therapy; selectin binding inhibitor; gene therapy; inflammation;

KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
polycystic kidney disease; polyomyositis; systemic sclerosis; dermatitis;

KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenocarcinoma;
Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;

KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
myocarditis; adult respiratory distress syndrome; eczema; psoriasis;

KW asthma; hypersensitivity; rheumatic fever; tissue rejection; ds.

OS Homo sapiens.

PN WO200106015-A1.

PD 25-JAN-2001.

PF 19-JUL-2000; 2000WO-US19741.

PR 20-JUL-1999; 99US-0144694.

PR 13-JUL-2000; 2000US-0593828.

PA (REGC) UNTV CALIFORNIA.

PI Rosen SD, Lee JK, Hemmerlich S;

DR WPI; 2001-138471/14.

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications -

PS Example 2; Page 116-123; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-6 (GST-6)

CC genomic DNA.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.
 CC XX
 SQ Sequence 27150 BP; 8357 A; 5396 C; 5398 G; 7998 T; 1 other:
 Query Match 1.3%; Score 18; DB 22; Length 27150;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1228 aataattgattctaaag 1245
 ||||||||||||||||
 DB 9623 aataattgattctaaag 9640
 RESULT 35
 AAF2497/C
 ID AAF2497 standard; CDNA; 81369 BP.
 AC AAF2497;
 XX
 DT 26-JAN-2001 (first entry)
 XX
 DE Human T gene DNA.
 KW T gene; human; central nervous system development; CNS; nootropic;
 KW neuroleptic; antidepressant; gene therapy; antisense; treatment;
 KW schizophrenia; autism; manic depression; mental retardation; ds.
 XX
 OS Homo sapiens.
 XX
 PN DEL19908423-A1.
 XX
 PD 31-AUG-2000.
 XX
 PF 26-FEB-1999; 99DE-1008423.
 XX
 PR 26-FEB-1999; 99DE-1008423.
 XX
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 XX
 PI Poustka A, Coy J;
 XX
 DR WPI; 2000-580150/55.
 XX
 DR P-PSDB; AAB10762.
 XX
 PT DNA encoding a protein involved in development of the central nervous
 PT system (CNS); antisense sequences, ribozymes and antibodies, useful for
 PT treatment of, e.g. schizophrenia and manic depression
 XX
 PS Claim 1a; Fig 2; 86pp; German.
 PS
 CC This invention describes a novel DNA sequence, which encodes a protein
 CC that is involved in development of the central nervous system (CNS) and
 CC has tissue and development-specific expression. The products of the
 CC invention have nootropic, neuroleptic and antidepressant activity and can
 CC be used for gene therapy and antisense inhibition. The method also
 CC describes a method for producing (1) antisense RNA that is complementary
 CC to DNA as above, which can reduce or inhibit synthesis of the protein

CC coding DNA; (2) a ribozyme, which is complementary to DNA as above, which
 CC specifically binds to and cleaves transcribed DNA, which reduces or
 CC inhibits synthesis of the protein coding DNA; (3) an expression vector,
 CC containing DNA as above, or which encodes antisense RNA or a ribozyme;
 CC (4) a host cell transformed with a vector as in (3); (5) a protein,
 CC encoded by DNA as above; (6) a method to produce the protein of (5)
 CC comprising culturing the cell of (4) and isolating the protein from the
 CC cell or the culture medium; (7) an antibody targeted against the protein
 CC of (5); (8) a diagnostic method to detect disturbed expression of the
 CC protein of (5) or to detect altered forms of the protein by contacting a
 CC sample with a DNA sequence or antibody and determining direct or indirect
 CC contact, and comparing the expression of the protein with a healthy
 CC patient; (9) a diagnostic kit to perform the method of (8); (10) a
 CC non-human transgenic animal, where the naturally occurring T gene has an
 CC altered gene structure or sequence; and (11) a method to produce a
 CC non-human animal as in (10). The DNA, derived from the T gene encodes a T
 CC protein (TP) which is involved in development of the central nervous
 CC system. Antisense sequences, ribozymes and antibodies are useful for
 CC treatment of disorders of the CNS including schizophrenia, autism, manic
 CC depression and mental retardation. This sequence encodes the human T
 CC protein described in the method of the invention.
 CC XX
 SQ Sequence 81369 BP; 26939 A; 14019 C; 13599 G; 26806 T; 6 other:

Query Match 1.3%; Score 18; DB 21; Length 81369;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 taataattgattctaa 1244
 ||||||||||||||||
 DB 76855 TAATAATTGATTCTAA 76838

RESULT 36
 AAF24497
 ID AAF24497 standard; CDNA; 240825 BP.
 AC AAF24497;
 XX
 DT 23-MAY-2001 (first entry)
 XX
 DE Human PG-3 gene.
 KW Human; PG-3; cancer; BRCA1; chromosome 8p23; ds.
 XX
 KM Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT misc_feature
 FT Location/Qualifiers
 FT 1..2000
 FT /note= "a
 FT 1823..1840
 FT /tag= b
 FT 1980..1998
 FT /tag= c
 FT 1987..2011
 FT /tag= d
 FT /note= " binds probe"
 FT replace(1999,C)
 FT /tag= e
 FT complement(2000..2018)
 FT /tag= f
 FT 2001..238825
 FT /tag= g
 FT /product= "PG-3"
 FT /note= "this sequence contains introns"
 FT 2001..2079
 FT /tag= h
 FT /label= "A"
 FT 2108..2125
 FT /tag= i
 FT 4559..4577
 FT primer_bind

```

FT primer_bind /tag= j
FT 4582..4600
FT /tag= k
FT misc_binding /tag= l
FT 4589..4613
FT /note= "binds probe"
FT replace(4601,G)
FT /tag= m
FT primer_bind complement(4602..4620)
FT /tag= n
FT exon 4627..4718
FT /tag= o
FT /label= "B"
FT primer_bind 4891..4908
FT /tag= p
FT primer_bind 10007..10025
FT /tag= q
FT 10115..10233
FT /tag= r
FT /label= "C"
FT primer_bind 10209..10227
FT /tag= s
FT 10216..10240
FT /tag= t
FT /note= "binds probe"
FT replace(10228,T)
FT /tag=
FT primer_bind complement(10229..10247)
FT /tag= v
FT primer_bind 10267..10285
FT /tag= w
FT 10274..10298
FT /tag= x
FT /note= "binds probe"
FT replace(10286,T)
FT /tag= y
FT primer_bind complement(10287..10305)
FT /tag= z
FT allele replace(10370,)
FT /tag= aa
FT 10411..10430
FT primer_bind /tag= ab
FT 26810..26897
FT exon /tag= ac
FT /label= "D"
FT 31357..31471
FT exon /tag= ad
FT /label= "E"
FT 34261..34404
FT exon /tag= ae
FT /label= "F"
FT 37377..37466
FT exon /tag= af
FT /label= "S"
FT 39704..40858
FT exon /tag= ag
FT /label= "T"
FT 39556..39574
FT primer_bind /tag= ah
FT 39877..39896
FT primer_bind /tag= ai
FT 39925..39943
FT primer_bind /tag= aj
FT 39932..39956
FT misc_binding /tag= ak
FT /note= "binds probe"
FT replace(39944,T)
FT /tag= al
FT allele complement(39945..39963)
FT /tag= am
FT primer_bind 39953..39970
FT /tag= an
FT primer_bind 39954..39972

FT misc_binding /tag= ao
FT 39961..39985
FT /tag= ap
FT /note= "binds probe"
FT replace(39973,C)
FT /tag= aq
FT primer_bind complement(39974..39992)
FT /tag= ar
FT primer_bind 40242..40259
FT /tag= as
FT primer_bind 41137..41154
FT /tag= at
FT primer_bind 41366..41384
FT /tag= au
FT misc_binding 41373..71397
FT /tag= av
FT /note= "binds probe"
FT replace(41385,C)
FT /tag= aw
FT 41385..41403
FT /tag= ax
FT primer_bind complement(41386..41404)
FT /tag= ay
FT 41392..41416
FT /tag= az
FT /note= "binds probe"
FT replace(41404,C)
FT /tag= ba
FT primer_bind complement(41405..41423)
FT /tag= bb
FT 41564..41581
FT /tag= bc
FT primer_bind 42122..42141
FT /tag= bd
FT 42213..42231
FT /tag= be
FT 42220..42244
FT /tag= bf
FT /note= "binds probe"
FT replace(42232,C)
FT /tag= bg
FT primer_bind complement(42233..42251)
FT /tag= bh
FT 42526..42543
FT primer_bind /tag= bi
FT 50436..50545
FT exon /tag= bj
FT /label= "G"
FT 67289..67309
FT primer_bind /tag= bk
FT 67456..67474
FT primer_bind /tag= bl
FT 67463..67487
FT /tag= bm
FT /note= "binds probe"
FT replace(67475,G)
FT /tag= bn
FT primer_bind complement(67476..67494)
FT /tag= bo
FT 67724..67741
FT primer_bind /tag= bp
FT 69182..69200
FT primer_bind /tag= bq
FT 69502..69520
FT primer_bind /tag= br
FT 69509..69533
FT misc_binding /tag= bs
FT /note= "binds probe"
FT replace(69521,G)
FT /tag= bt
FT allele complement(69522..69540)
FT /tag= bu
FT primer_bind 69609..69626

```

```
FT primer_bind /tag= bv
FT 72698..72715 /tag= bw
FT primer_bind /tag= bw
FT 72819..72837 /tag= bx
FT misc_binding /tag= bx
FT 72826..72850 /tag= by
FT allele /note= " binds probe"
FT replace(72838,T)
FT /tag= bz
FT primer_bind Complement(72839..72857)
FT /tag= ca
FT exon 72881..72918
FT /tag= cb
FT /label= "H"
FT 73099..73117
FT /tag= cc
```

```
Query Match 1.3%; Score 18; DB 22; Length 240825;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1207 aaagatgacatgtgtt 1224
Db 90632 aaagatgacatgtgtt 90649
```

```
RESULT 37
AA161371
ID AA161371 standard; DNA; 335913 BP.
XX
AC AA161371;
XX
DT 16-OCT-2001 (first entry)
XX
DE Soybean 240017 region G3, SEQ ID NO: 2.
XX
KW Soybean; anthelmintic; gene therapy; soybean cyst nematode; SCN;
KW SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW 240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS Glycine max.
XX
PN WO200151627-A2.
XX
PD 19-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-US00552.
XX
PR 07-JAN-2000; 2000US-0174880.
XX
PA (MONS ) MONSANTO CO.
XX
PI Hauge BM, Wang ML, Parsons JD, Parnell LD;
XX
DR WPI: 2001-425872/45.
XX
DR P-PSDB: AAM42214.
XX
PT New purified nucleic acid for producing a soybean plant having soybean
PT cyst nematode resistance and for use in plant breeding programs -
XX
PS Claim 2; Page 204-400; 1353pp; English.
XX
XX The invention relates to nucleic acid molecules from regions of the
CC soybean genome which are associated with soybean cyst nematode (SCN)
CC resistance. The nucleic acids are used to transform plants, and can
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC of soybean plants and for introgressing SCN resistance or partial SCN
CC resistance into soybean plants. They can also be used in plant breeding
CC programmes. The invention also relates to proteins encoded by such
CC nucleic acid molecules, as well as antibodies capable of recognising
CC these proteins. The present sequence is a nucleic acid molecule
```

```
CC provided in the specification.
XX
SQ Sequence 335913 BP; 114579 A; 53403 C; 53026 G; 114905 T; 0 other;
```

```
Query Match 1.3%; Score 18; DB 22; Length 335913;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 293 tggagatgatgatgtcg 310
Db 125725 tggagatgatgatgtcg 125742
```

```
RESULT 38
AA161372
ID AA161372 standard; DNA; 335913 BP.
XX
```

```
AC AA161372;
XX
DT 16-OCT-2001 (first entry)
XX
```

```
DE Soybean 240017 region G3, SEQ ID NO: 3.
```

```
XX
KW Soybean; anthelmintic; gene therapy; soybean cyst nematode; SCN;
KW SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW 240017 region G3; 318013 region A3; 515002 region G2; ds.
```

```
XX
OS Glycine max.
```

```
XX
PN WO200151627-A2.
```

```
XX
PD 19-JUL-2001.
```

```
XX
PF 05-JAN-2001; 2001WO-US00552.
```

```
XX
PR 07-JAN-2000; 2000US-0174880.
```

```
XX
PA (MONS ) MONSANTO CO.
```

```
XX
PI Hauge BM, Wang ML, Parsons JD, Parnell LD;
```

```
XX
DR WPI: 2001-425872/45.
XX
DR P-PSDB: AAM42215.
```

```
XX
PT New purified nucleic acid for producing a soybean plant having soybean
PT cyst nematode resistance and for use in plant breeding programs -
```

```
XX
PS Claim 2; Page 400-595; 1353pp; English.
```

```
XX The invention relates to nucleic acid molecules from regions of the
CC soybean genome which are associated with soybean cyst nematode (SCN)
CC resistance. The nucleic acids are used to transform plants, and can
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC of soybean plants and for introgressing SCN resistance or partial SCN
CC resistance into soybean plants. They can also be used in plant breeding
CC programmes. The invention also relates to proteins encoded by such
CC nucleic acid molecules, as well as antibodies capable of recognising
CC these proteins. The present sequence is a nucleic acid molecule
CC provided in the specification.
```

```
XX
SQ Sequence 335913 BP; 114582 A; 53398 C; 53027 G; 114906 T; 0 other;
```

```
Query Match 1.3%; Score 18; DB 22; Length 335913;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 293 tggagatgatgatgtcg 310
Db 125725 tggagatgatgatgtcg 125742
```

```
RESULT 39
AAF21610
ID AAF21610 standard; DNA: 349980 BP.
XX
AC AAF21610;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
XX
PN WO200066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000MO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99MO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Piazza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisseria infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisseria bacteria or as a diagnostic reagent for detecting the
CC presence of Neisseria bacteria or of antibodies raised to Neisseria
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;
```

```
Query Match 1.3%; Score 18; DB 21; Length 349980;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 49 ttatctctgttttaacc 66
|||||
```

DB 137432 ttatctctgttttaacc 137449

```
RESULT 40
AAV21209
ID AAV21209 standard; DNA: 1664976 BP.
XX
AC AAV21209;
XX
DT 10-NOV-1998 (first entry)
XX
DE Methanococcus jannaschii circular chromosome.
XX
KM Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KM genome; autotrophic; extrachromosomal element; identification; ds.
XX
OS Methanococcus jannaschii.
XX
PN WO9807830-A2.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97MO-US14900.
XX
PR 22-AUG-1996; 96US-0024428.
XX
PA (GENO-) INST GENOMIC RES.
PA (UNIT) UNIV ILLINOIS FOUND.
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;
XX
DR WPI; 1998-169145/15.
XX
PT Complete genome sequence of methano-genic archaeon, Methanococcus
PT jannaschii - useful in identification of M. jannaschii genome
PT fragment
XX
PS Claim 13; Page 152-585; 614pp; English.
XX
CC The present sequence represents the complete 1.66-megabase pair genome
CC sequence of the Methanococcus jannaschii circular chromosome. The
CC present invention describes M. jannaschii open reading frames from the
CC genome sequence. The invention also describes a computer based system
CC for identifying fragments of the M. jannaschii genome that are
CC homologous to target nucleotide sequences, comprising: (a) data storage
CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550
CC bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide
CC sequence at least 99.9% identical to it; (b) search means for comparing a
CC target sequence to the nucleotide sequence of the data storage means to
CC identify a homologous sequence, and (c) retrieval means for obtaining
CC the homologous sequence. The method, which is based on whole genome
CC random sequencing of an autotrophic archaeon M. jannaschii, the genome
CC of which consists of 3 physically distinct elements, a large circular
CC chromosome (the 1664976 bp sequence given in AAV21209), a large circular
CC extra-chromosomal element (the 58407 bp sequence given in AAV21210), and
CC a small circular extra-chromosomal element (the 16550 bp sequence given
CC in AAV21211), can be used in the identification of M. jannaschii genome
CC fragment.
XX
SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;
```

```
Query Match 1.3%; Score 18; DB 19; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 721 tcagagttgaaaaaac 738
|||||
DB 146412 tcagagttgaaaaaac 146429
```

RESULT 41


```

XX OS Homo sapiens.
XX PN WO200016916-A2.
XX PD 06-APR-2000.
XX PF 23-SEP-1999; 99WO-US222226.
XX PR 28-SEP-1998; 98US-0102161.
XX PR 28-SEP-1998; 98US-0102180.
XX PR 29-SEP-1998; 98US-0102380.
XX PR 08-OCT-1998; 98US-0103815.
XX PR 27-OCT-1998; 98US-0105877.
XX PA (CHIR ) CHIRON CORP.
XX PA (HYSE-) HYSEQ INC.
XX PI Williams LT, Escobedo J, Tunis MA, Garcia PD, Sudduth-Klinger J;
XX PI Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX PI Lamson G, Drmanac R, Civenjakov R, Dickson M, Drmanac S, Labat I;
XX PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX PS Claim 1; Page 344; 502pp; English.
XX PT Polynucleotide library comprising 1079 defined sequences, useful in
XX PT the form of an array to detect cancer or susceptibility to cancer -
XX PS Claim 1; Page 344; 502pp; English.
XX CC The present invention describes a library of polynucleotides comprising
XX CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
XX CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
XX CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
XX CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that
XX CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
XX CC of detecting differentially expressed genes correlated with a cancerous
XX CC state of a mammalian cell comprising detecting a gene product encoded by
XX CC one of the 1079 sequences given in the specification. The polynucleotides
XX CC are used to monitor patients having (or susceptible) to cancer to detect
XX CC potentially malignant events at a molecular level before they are
XX CC detectable at a gross morphological level. The polynucleotides are also
XX CC useful for monitoring the efficacy of various therapies and preventive
XX CC interventions. Polynucleotide probes based on the disclosed sequences
XX CC are useful for chromosome mapping and detection of transcription levels.
XX CC The 1079 polynucleotide sequences were derived from a human colon cancer
XX CC cell line Km12cl-A cDNA library.
XX SQ Sequence 376 BP; 94 A; 90 C; 113 G; 79 T; 0 other;
SQ Query Match 1.2%; Score 17; DB 21; Length 376;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 755 tggagacggtgaacc 771
DB 104 TGGAGACGCGTGAACCC 88
RESULT 44
AAH35461
ID AAH35461 standard; cDNA; 553 BP.
AC AAH35461;
XX 03-SEP-2001 (first entry)
XX DE Human colon cancer antigen encoding cDNA SEQ ID NO:2543.
XX KW Human colon cancer; colon cancer antigen; diagnosis: detection;
XX KW colorectal carcinoma; chromosome 17; ss.
XX

```

```

OS OS Homo sapiens.
XX PN WO200122920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMA GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX PI WPI: 2001-235357/24.
XX DR P-PSDB; AAG76056.
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX PS Claim 1; Page 4128-4132; 9803pp; English.
XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where
XX CC the proteins are collectively known as colon cancer antigens. The colon
XX CC cancer antigens have cytostatic activity and can be used in gene
XX CC therapy and vaccine production. N and P may be used in the prevention,
XX CC diagnosis and treatment of diseases associated with inappropriate P
XX CC expression. For example, N and P may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of P by expressing
XX CC inactive proteins or to supplement the patient's own production of P.
XX CC Additionally, N may be used to produce the colon cancer-associated P,
XX CC by inserting the nucleic acids into a host cell and culturing the cell
XX CC to express the proteins. N and P can be used in the prevention, diagnosis
XX CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX CC and AAG77789 represent sequences used in the exemplification of the
XX CC present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX CC missing at time of publication, meaning no sequences are present for
XX CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX SQ Sequence 553 BP; 155 A; 108 C; 134 G; 150 T; 6 other;
SQ Query Match 1.2%; Score 17; DB 22; Length 553;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 46 gtttcatcctgtttt 62
DB 130 gtttcatcctgtttt 146
RESULT 45
AAI43142
ID AAI43142 standard; DNA; 575 BP.
AC AAI43142;
XX 17-OCT-2001 (first entry)
XX DE Probe #11828 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX

```

```

XX 30-JAN-2001; 2001WC-US00663.
PF
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 11828; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 575 BP; 179 A; 87 C; 79 G; 230 T; 0 other;

Query Match          1.2%; Score 17; DB 22; Length 575;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 553 aatatttcattgtcttc 569
   |||||||
Db 24 aatatttcattgtcttc 40

RESULT 46
AAH61796/c
ID AAH61796 standard; DNA; 724 BP.
XX
AC AAH61796;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein coding sequence, t4-15.nt.
XX
KM Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX
OS Borrelia burgdorferi.
XX
PN W09859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI (MED-) MEDIMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI; 1999-189980/16.
DR P-PDB; AAY20099.
XX

```

```

PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 1; Page 196-197; 275pp; English.
XX
CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 724 BP; 347 A; 125 C; 85 G; 167 T; 0 other;

Query Match          1.2%; Score 17; DB 20; Length 724;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 tgtgtggcgaagatttc 693
   |||||||
Db 294 tctgtgcgcacagatttt 278

RESULT 47
AAH67794
ID AAH67794 standard; DNA; 732 BP.
XX
AC AAH67794;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 2823.
XX
KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX
KM organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0158162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senon A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PDB; AAG92575.
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
PS Claim 8; SEQ ID NO: 2823; 246bp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing

```

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX
 SQ Sequence 732 BP; 168 A; 201 C; 197 G; 166 T; 0 other;

Query Match 1.2%; Score 17; DB 22; Length 732;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1273 gcagaaagcttgaaga 1289
 |||||
 DB 145 gcagaaagcttgaaga 161

RESULT 48
 AAX36834/C
 ID AAX36834 standard; DNA; 816 BP.

XX
 AC AAX36834;

XX
 DT 14-JUL-1999 (first entry)

XX
 DE Human XLIS gene fragment including exon 4.

XX
 KW XLIS gene; human; detection; diagnosis; prenatal diagnosis; therapy;
 KW lissencephaly; LIS; agyria-pachygyria; subcortical laminar heterotopia;
 KW SCH; cortical dysgenesis; cryptogenic epilepsy; neurological disorder;
 KW neurodegenerative disease; Alzheimer's disease; X-linked disorder;
 KW genetic counselling; ss.

XX
 OS Homo sapiens.

XX
 PN EP918091-A1.

XX
 PD 26-MAY-1999.

XX
 PF 21-NOV-1997; 97EP-0402811.

XX
 PR 21-NOV-1997; 97EP-0402811.

XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX
 PI Chelly J, Des Portes V, Kahn A, Pinard J;

XX
 DR WPI; 1999-290318/25.

PT New gene and its gene product expressed in the brain, useful for
 PT diagnosing and treating disorders such as lissencephaly and
 PT subcortical laminar heterotopia

XX
 PS Claim 1; Page 23; 71pp; English.

XX This sequence is a fragment of the human XLIS gene of the invention.
 CC The XLIS fragments may be used to detect abnormalities in the expression
 CC of the XLIS gene transcripts or to compare their sequence with that of
 CC the XLIS transcripts from patients for in vitro especially prenatal
 CC diagnosis of lissencephaly (LIS) (or agyria-pachygyria), subcortical
 CC laminar heterotopia (SCH), cortical dysgenesis, cryptogenic epilepsies
 CC or neurodegenerative diseases such as Alzheimer's disease. These
 CC disorders mainly affect females as the XLIS gene is X-linked. These
 CC fragments may also be used to administer to patients to prevent or treat
 CC the above disorders and may be used as a tool in genetic counselling.
 CC Oligonucleotides which bind to the fragments may be used to amplify the
 CC XLIS gene from a sample for comparison to normal samples in the in vitro
 CC diagnosis regime. This may also be performed by amplifying XLIS cDNA from
 CC the mRNA in the sample. Antibodies to XLIS may be used to detect XLIS in
 CC a biological sample or can be administered to patients to prevent or
 CC treat the above disorders. They may also be used to purify XLIS from a

CC biological sample. XLIS may also be administered to patients to prevent
 CC or treat the above neurological disorders. In addition XLIS may be used
 CC as a marker of neuronal cells at an early stage of development; its
 CC discovery increases understanding of both the neuronal movement which
 CC leads to development of the cortical region of the brain and of the
 CC pathogenesis of the group of neuronal disorders mentioned above.

XX
 SQ Sequence 816 BP; 204 A; 170 C; 178 G; 262 T; 2 other;

Query Match 1.2%; Score 17; DB 20; Length 816;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 agagaaagctccca 396
 |||||
 DB 353 AGGAGAAAGCTCCCA 337

RESULT 49
 AAX61795/C
 ID AAX61795 standard; DNA; 837 BP.

XX
 AC AAX61795;

XX
 DT 19-JUL-1999 (first entry)

XX
 DE B. burgdorferi antigenic protein coding sequence, f4-15.nt.

XX
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

XX
 OS Borrelia burgdorferi.

XX
 PN WO9859071-A1.

XX
 PD 30-DEC-1998.

XX
 PF 18-JUN-1998; 98WO-US12718.

XX
 PR 03-SEP-1997; 97US-0057483.

XX
 PR 20-JUN-1997; 97US-0050359.

XX
 PR 22-JUL-1997; 97US-0053344.

XX
 PR 22-JUL-1997; 97US-0053377.

XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX
 PI (MEDI-) MEDIMUNE INC.

XX
 PI Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX
 DR WPI; 1999-18980/16.

XX
 DR P-PSDB; AAY20098.

PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of diseases
 PT caused by Borrelia, particularly Lyme disease

XX
 PS Claim 1; Page 196; 275pp; English.

XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus.

XX
 SQ Sequence 837 BP; 396 A; 134 C; 98 G; 209 T; 0 other;

Query Match 1.2%; Score 17; DB 20; Length 837;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 tctgtggcaagattt 693
 Db 344 TGTGTGGCAAGATT 328

Db 336 ctaataattgattct 352

RESULT 50

AAC59007
 ID AAC59007 standard; cDNA: 917 BP.

AC AAC59007;

DT 02-FEB-2001 (first entry)

DE Human secreted protein coding sequence SEQ ID NO: 28.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KM antiallergic; hepatotropic; antidiabetic; antineoplastic; antitumor;
 KM vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KM neurological disease; infection; human; secreted protein; ss.

OS Homo sapiens.

PN WO200055175-A1.

PD 21-SEP-2000.

PF 09-MAR-2000; 2000WO-US06049.

PR 12-MAR-1999; 99US-0124144.

PR 11-JUN-1999; 99US-0138574.

PR 03-DEC-1999; 99US-0168667.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis G;

DR WPI; 2000-638175/61.

DR P-PDB; AAB27577.

PS Claim 1; Page 353; 428pp; English.

XX The invention relates to the isolation of genes AA58990-A59039 encoding
 CC 50 human secreted proteins AAB27560-B27609. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (SEQID) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.

SO Sequence 917 BP; 281 A; 170 C; 166 G; 300 T; 0 other;

Query Match 1.2%; Score 17; DB 21; Length 917;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: May 2, 2002, 13:20:48
 Job time: 6189 sec

QY 1226 ctaataattgattct 1242

Fri May 3 10:57:49 2002

us-09-645-192-1.01110.rng

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2002, 07:39:39 ; Search time 24.66 Seconds

(Without alignments)
1349.767 Million cell updates/sec

Title: US-09-645-192-2

Perfect score: 2389

Sequence: 1 MKIFKCYFKHTLQKVFILF.....DWTLPEKLFMDRNLTITS 453

Scoring table:

BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_1101.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2389	100.0	453	22	AAUD0037 Human C2Gnt3. Hom
2	868	36.3	438	21	AAAB18995 A core 2 beta-1,6-
3	868	36.3	438	21	AAAY94992 Human C2/Agnt prot
4	868	36.3	438	22	AAAB30518 Amino acid sequenc
5	868	36.3	465	22	AAAG75058 Human colon cancer
6	868	36.3	663	21	AAAB18996 A partial core 2 b
7	865.5	36.3	428	19	AAAB53942 Rat DHI protein.
8	866	36.2	465	21	AAAB53444 Human pancreatic c
9	863.5	36.1	428	19	AAAB39443 Human core 2 Gnt p
10	863.5	36.1	428	21	AAAB30298 Human heart core 2
11	862	36.1	406	21	AAAB18999 A core 2 beta-1,6-

12	854.5	35.8	428	15	AAAB51386 Sequence of human
13	850	35.6	427	21	AAAB30297 Diabetic rat heart
14	711	29.8	400	16	AAAB71932 I-branched enzyme
15	711	29.8	400	17	AAAR92474 Beta-1,6-N-acetyl
16	711	29.8	400	19	AAAB56628 Full length amino
17	711	29.8	400	19	AAAB47184 Human beta-1,6-N-a
18	620	26.0	120	22	AAAB24019 Human EST encoded
19	383	16.0	126	17	AAAB92475 C2Gnt catalytic do
20	355	14.9	126	19	AAAB56627 Human beta-1,6-N-a
21	355	14.9	161	19	AAAB47185 Human acid sequenc
22	224.5	9.4	864	22	AAAB62912 Amino acid sequenc
23	219.5	8.9	865	22	AAAB62910 Amino acid sequenc
24	213.5	8.9	821	22	AAAB62911 Amino acid sequenc
25	213.5	8.9	837	22	AAAB62909 Amino acid sequenc
26	172	7.2	64	22	AAAB30519 Murine beta-1,6-N-
27	166.5	7.0	426	21	AAAB90938 Arabidopsis thalia
28	166.5	7.0	434	21	AAAB90937 Arabidopsis thalia
29	158	6.6	334	21	AAAB90939 Arabidopsis thalia
30	151	6.3	447	21	AAAB43728 Arabidopsis thalia
31	148	6.2	486	21	AAAB28142 Arabidopsis thalia
32	132	5.5	484	21	AAAB31854 Arabidopsis thalia
33	132	5.5	630	21	AAAB31853 Arabidopsis thalia
34	119	5.0	316	18	AAAB14080 S.thermophilus exo
35	119	5.0	316	18	AAAB22179 S.thermophilus exo
36	115	4.8	348	21	AAAB28143 Arabidopsis thalia
37	113.5	4.8	794	22	AAAB0426 P. falciparum telo
38	109	4.6	74	22	AAAB15758 Peptide #2192 enco
39	109	4.6	74	22	AAAB28267 Peptide #2304 enco
40	109	4.6	74	22	AAAB03492 Peptide #2174 enco
41	109	4.6	289	21	AAAB43729 Arabidopsis thalia
42	108	4.5	165	21	AAAB15071 Arabidopsis thalia
43	108	4.5	754	22	AAAB82101 S. epidermidis ope
44	107	4.5	358	21	AAAB28144 Arabidopsis thalia
45	106	4.4	33	17	AAAB92476 C2Gnt C-terminal r

ALIGNMENTS

RESULT 1	AAUD0037 standard; Protein; 453 AA.
XX	AAUD0037:
AC	14-MAY-2001 (first entry)
XX	Human C2Gnt3.
DE	Human: C2Gnt3. Thymus-related disorder; cancer; tumour; adenoma;
KW	UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GalNAc transferase; sarcoma;
KW	malignant melanoma; breast cancer; cervical cancer; hypoxanthyl;
KW	hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
KW	leukaemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;
KW	acquired immunodeficiency syndrome; sepsis; wound healing; infection.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Domain
FT	Protein
FT	Location/Qualifiers
FT	12..32
FT	/label= "Transmembrane domain"
FT	39..453
FT	/label= "protein fragment retaining enzyme activity"
FT	/note= "used in a fusion protein for in vitro
FT	expression studies"
FT	Modified-site
FT	72
FT	/note= "N-glycosylated"
FT	Modified-site
FT	286
FT	/note= "N-glycosylated"
FT	Modified-site
FT	317
FT	/note= "N-glycosylated"
FT	Modified-site
FT	448
FT	/note= "N-glycosylated"

XX WO200114535-A2.
 XX 01-MAR-2001.
 XX 24-AUG-2000; 2000WO-DK00469.
 XX 24-AUG-1999; 99US-0150488.
 XX (SCHW/) SCHWIENTEK T.
 XX (CLAU/) CLAUSEN H.
 XX Schwiientek T, Clausen H;
 XX WPI: 2001-226615/23.
 XX N-PSDB; AAS00043.
 XX New C2Gnt3 polypeptides and nucleic acids encoding the polypeptides
 XX useful for treating conditions mediated by a C2Gnt3 polypeptide, e.g.,
 XX thymus-related disorders, cancers, tumours, immunosuppression
 XX
 XX Claim 17; Fig 1; 97pp: English.
 XX
 XX The sequence represents Human UDP-N-acetyl-glucosamine:Galactose-beta1,
 XX 3-N-acetylglactosamine-alpha-R beta1-6-N-acetylglucosaminyltransferase
 XX (UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
 XX C2Gnt3). C2Gnt3 and nucleic acids encoding it are useful in the
 XX preparation of compositions for treating a conditions mediated by C2Gnt3,
 XX particularly a thymus-related disorder. C2Gnt3, nucleic acids
 XX encoding it and antibodies against it may also be used for in vitro
 XX purposes related to scientific research, DNA synthesis and manufacture of
 XX vectors, in the prognostic and diagnostic evaluation of conditions
 XX associated with altered expression or activity of C2Gnt3 or conditions
 XX requiring modulation of C2Gnt3, as well as in monitoring conditions by
 XX detecting and localising the DNA and protein. Disorders such as tumours
 XX (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
 XX breast or cervix), hypocoactivity, hyperactivity, atrophy, enlargement of
 XX thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
 XX acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
 XX sepsis, wound healing, acute and chronic infection, cell-mediated or
 XX humoral immunity, or TH1/TH2 imbalance, may be treated using these protein
 XX or nucleic acid. The antibodies may be used to screen potential
 XX therapeutic compounds to determine their effects on a conditions such as
 XX thymus-related disorder or cancer, to determine the level of C2Gnt3
 XX expression in cells genetically engineered to produce C2Gnt3, or to
 XX detect and quantify polypeptides in a sample to determine their role in a
 XX particular cellular events or pathological states and to diagnose and
 XX treat such pathological states.
 XX
 XX Sequence 453 AA;
 XX
 XX Query Match 100.0%; Score 2389; DB 22; Length 453;
 XX Best Local Similarity 100.0%; Pred. No. 2.1e-207;
 XX Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX 1 MKTFKCYFKHLLQOKYFLFLTMTLSLKLVNRRFLPOKDIYEVLSLSPFRNRY 60
 XX 1 MKTFKCYFKHLLQOKYFLFLTMTLSLKLVNRRFLPOKDIYEVLSLSPFRNRY 60
 XX 61 THAKDEVREVNCSGIYEDPELIGKSLERRRDIIDLEDDDVAVMTSDCDIYQTLRGYA 120
 XX 61 THAKDEVREVNCSGIYEDPELIGKSLERRRDIIDLEDDDVAVMTSDCDIYQTLRGYA 120
 XX 61 THKDEVREVNCSGIYEDPELIGKSLERRRDIIDLEDDDVAVMTSDCDIYQTLRGYA 120
 XX 121 QXIVSKEEKSFRTAYSLVYHKDAIMVERLIHATYNOHNHYCHYDKKADPTKRVANNNA 180
 XX 121 QXIVSKEEKSFRTAYSLVYHKDAIMVERLIHATYNOHNHYCHYDKKADPTKRVANNNA 180
 XX 121 QXIVSKEEKSFRTAYSLVYHKDAIMVERLIHATYNOHNHYCHYDKKADPTKRVANNNA 180
 XX 181 KCSNFIASKLEAVENAHISRLQADUNCLSDLLKSSIQMKYVYINLCGDFPLKSNFELY 240
 XX 181 KCSNFIASKLEAVENAHISRLQADUNCLSDLLKSSIQMKYVYINLCGDFPLKSNFELY 240
 XX 181 KCSNFIASKLEAVENAHISRLQADUNCLSDLLKSSIQMKYVYINLCGDFPLKSNFELY 240
 XX 241 SELKTLNGAMMLETVKRPNSKLERFTYHHELRVPEYVYKLPRTNISKEAPHNIGIV 300

DB 241 selktlngammlenvkxpnskierfthelrrvpeyvkpirtnskeapnngitv 300
 QY 301 GSAYFVLQAFVKYIFNNNSIVQDFFAWMSKDTYSPDEHFHATLLRVSPIDPEISRSADVS 360
 DB 301 gsayfvlqsatfvkyilfnnsivqdfawmskdtyspdehfhwatlrvspidpeisrsadvs 360
 QY 361 DLOSKTRLVKWNYYEGFRFPYECTSGHRSVCITVGAALRWLIKDGHHFAKFSKVPIL 420
 DB 361 dlosktrlvkwnyyegfrfpyectshirsvciygaaelrwlkdgghfankfskvdpl 420
 QY 421 IKCLAERLEEOQPDMLTPSEKLFMDRLNLTTS 453
 DB 421 ikclaerleegqrdwiltpekslmdrnlntts 453
 XX
 XX RESULT 2
 XX AAB18995
 XX ID AAB18995 standard; Protein: 438 AA.
 XX
 XX AAB18995;
 XX
 XX 08-FEB-2001 (first entry)
 XX
 XX A core 2 beta-1,6-N-acetylglucosaminyltransferase.
 XX
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 XX cancer; cardiovascular disorder; inflammatory disorder; asthma;
 XX rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 XX septic shock; adult respiratory distress syndrome; ARDS; cancer;
 XX platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
 XX clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 XX diverticulitis; ulcerative colitis.
 XX
 XX Homo sapiens.
 XX
 XX CA2296936-A1.
 XX
 XX 03-AUG-2000.
 XX
 XX 03-FEB-2000; 2000CA-2296936.
 XX
 XX 03-FEB-1999; 99US-0118674.
 XX
 XX (GLYC-) GLYCDESIGN INC.
 XX
 XX Korczak B, Lew A;
 XX
 XX WPI: 2000-594746/57.
 XX N-PSDB; AAA96569.
 XX
 XX New nucleic acid molecules of core 2
 XX beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 XX compositions for treatment of disorders mediated by the enzyme
 XX including cancer, cardiovascular and inflammatory disorders.
 XX
 XX Claim 3; Page 50-51; 66pp: English.
 XX
 XX The present sequence represents a human core 2
 XX beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 XX The polypeptide can be used to treat diseases and disorders, such as
 XX cancer, cardiovascular disorders and inflammatory disorders including
 XX asthma, rheumatoid arthritis, inflammatory bowel disease,
 XX arteriosclerosis, septic shock, adult respiratory distress syndrome
 XX (ARDS) and cancer. Various platelet-mediated pathologies such as
 XX arteriosclerosis and clotting can also be treated. The polypeptides of
 XX the invention are predominantly expressed in gastrointestinal tissue
 XX (stomach, colon, intestine, testis) and are elevated in cancer.
 XX Gastrointestinal disorders that may be prevented or treated include
 XX ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 XX and ulcerative colitis. The antibodies may be used in
 XX immuno-chemical analysis, to detect the novel polypeptide and to
 XX localize it to particular cells and tissues and to specific subcellular

DT 06-MAR-2001 (first entry)
 XX Amino acid sequence of beta-1-6-N-acetylglucosaminyltransferase.
 DE Human: beta-1-6-N-acetylglucosaminyltransferase; C2GNT-M: inflammation;
 XX membrane protein; branched sialyl lex; L-selectin; immune reaction;
 KW Inflammation; tissue rejection; tumour metastasis.
 KM
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 289
 FT /note= "potential N-glycosylation site"
 XX
 XX US6136580-A.
 XX
 XX 24-OCT-2000.
 XX
 XX 19-JAN-1999; 99US-0233506.
 XX
 XX 19-JAN-1999; 99US-0233506.
 XX
 XX (BURN-) BURNHAM INST.
 XX
 XX Fukuda M, Yeh J;
 XX
 XX WPI; 2001-040238/05.
 DR N-PSDB; AAC62134.
 XX
 XX
 PT New C2GNT-M polypeptides having core 2, core 4 and I branching
 PT beta-1-6-N-acetylglucosaminyltransferase activities for preparing
 PT reagents useful for diagnosing, preventing or treating inflammation or
 PT tumour metastasis.
 PT
 XX
 XX Claim 1; Fig 4; 25pp; English.
 PS
 PS
 CC The present sequence represents a human
 CC beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4
 CC and I branching activities. It is designated C2GNT-M. C2GNT-M is a
 CC membrane protein that is predominantly expressed in colon, small
 CC intestine, trachea, stomach and thyroid, as well as in certain cancer
 CC cell lines. C2GNT-M polypeptides may be used to prepare molecules having
 CC highly branched sialyl lex and L-selectins, which may be subsequently
 CC used to modulate immune reactions, e.g. inflammation and tissue
 CC rejection, and to prevent or inhibit tumour metastasis.
 CC
 XX
 XX Sequence 438 AA;
 SQ
 Query Match 36.3%; Score 868; DB 22; Length 438;
 Best Local Similarity 43.0%; Pred. No. 5 6e-70;
 Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;
 QY 23 LWLISLILKILNV-----RRLFPQKDIYVEYLSLSPFYRN-RYTHVDEVREYVNCSG 75
 Db 13 lwaigcmilatvalklsfrlkcdshlglesresqgycrnlllyflkxprksincsg 72
 QY 76 IY--EQLPL--EIGKSLERIRROIIDLEDDVYAMTSDCQIYOTLGGVQKUSKEKPSF 131
 Db 73 vtrgdgaavgalnmlevkkr-epftdchysiltdechkaekifgfpjskeevf 131
 QY 132 PIASLVVHRDAIWEPLIAITNQHNIYCIHYDRKAPDTFKVAMNNLAKCFSNIFIAK 191
 Db 132 playsmvlhekienferllraayapgnlycvhvdkspefkcaavkalisccfnvflask 191
 QY 192 LEAVEYAHISRLQDLCLDKSSIQMKYVINLGGDPLKSNFELVSELKILGANV 251
 Db 192 lrvvyvswstrvgadlncmedllsgsvpkryflintcgttdpiksnemvagalmlngins 251
 QY 252 LETVAPNSKLEPFYTHELRRVPEYEVKLPITNISKEAPPHNIQFVGSAYFVLGQAF 311
 Db 252 mesevppkhketrwkyhfevfr--dlhl--lnkkkdpypnllmftgnaylvasrdt 305

QY 312 VKIIFNNSIVODFFAMSKDQTSPEDEHFMATLIRVGPISGEI-SRSADVDSDLOKTRLYK 370
 Db 306 vqhvltkpkssqlllewxvdytspdehlwatlqarmpgsvpnkxydismsiatlvk 365
 QY 371 WNYEGEPF-----YPSCTGSHLRSCVCIYGAALRWLIRKDHWFANKEDSKVDPIILIKCLA 425
 Db 366 wqhgagdldkgyapcsglnqraicvygagdlmwlmqnhlllankfdpkvcdnalqcle 425
 QY 426 EKL 428
 Db 426 eyl 428
 RESULT 5
 ID AAG75058
 XX AAG75058 standard; Protein: 465 AA.
 AC AAG75058;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen protein SEQ ID NO:5822.
 XX
 KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 15.
 XX
 OS Homo sapiens.
 OS
 PM WO200122920-A2.
 PD 05-APR-2001.
 XX
 XX 28-SEP-2000; 2000WO-US26524.
 PE
 XX 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;
 PI WPI; 2001-235357/24.
 DR N-PSDB; AAH34463.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 PT
 XX
 XX Claim 11; Page 7338-7340; 9803pp; English.
 PS
 PS AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 XX Sequence 465 AA;
 SQ
 Query Match 36.3%; Score 868; DB 22; Length 465;

Best Local Similarity 43.0%; Pred. No. 6.1e-70;
Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;

QY 23 LMLSLKLKLVN-----RLFPQKDIYVEYSLSSTSPFVRN-RYTHVKDEVREYVNCSG 75
DB 40 lwaigcymllatvalklsfirkcdshlglesresgsqycrnllnflkpkarsincsg 99
QY 76 IV-EQEPPL-EIGKSLERIRRDIDLEDDDVAMTSDCDIYQTLRGYAQKIVSKREKSF 131
DB 100 vtrgdqeaavlqallnmlvkkrr-epftdthylsltrdcenhfaekrfiqfplskeevf 158
QY 132 PIATSLVYHKDAIMVERLHAIVNQHNICIHDKRAKAPTEFKVAMNNLAKCFSNIFTASK 191
DB 159 playsmvlnhekienferlliravypqnlcyvhvdekspetfkeavvkaaiscfpnvflask 218
QY 192 LEAVEYAHISRLQADNLCSLDLSSSIOMKYVINLGGDFPKKSNELSELKUNGAM 251
DB 219 lrvvvaswstrvgadlnmedllqssvwpkyflntcgtfdipknaemvqalkmlngns 278
QY 252 LETVPPNSKLEFRFYHHELRVRYEVYKLPFRINISKAAPPHNIQIFVGSAYVLSQAF 311
DB 279 mesevppkhketrwkyhfevr---dtlhl---tnkkkdpppylnlmtfgnayivasrdf 332
QY 312 VKYIFNNSIVQDFEAFMSKDTYSPDEHFATLIRVPGIGEI-SRSADVSDLSQKTRLVK 370
DB 333 vqhvlnknpksqqlievwkdtyspdehlwatlqratwmpsvpnphkpydisdmtslarlvk 392
QY 371 WNYEGFF-----YPSCTGSHLRSVCYGAELRMLIKDGHWFANKFDSKVDPIILKCLA 425
DB 393 wqhggedldkgapyapcsghnqratlcvygagdlmwlnqnhllanlkfddkvdnalqcle 452
QY 426 EKL 428
DB 453 eyl 455

RESULT 6
ID AAB18996 standard; Protein: 663 AA.

AC AAB18996;
DT 08-FEB-2001 (first entry)
XX

DE A partial core 2 beta-1,6-N-acetylglucosaminyltransferase.
XX
XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis.

XX
XX Homo sapiens.
OS
XX CA2296936-A1.
PN
XX
XX 03-AUG-2000.
PD
XX
XX 03-FEB-2000; 2000CA-2296936.
PF
XX
XX 03-FEB-1999; 99US-0118674.
PR
XX (GLYC-) GLYCDESIGN INC.
PA
XX
XX Koczak B, Lew A;
PI
XX
XX WPI, 2000-594746/57.
DR
XX
XX New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme

PT including cancer, cardiovascular and inflammatory disorders.
XX
XX Claim 3; Page 52; 66pp; English.

The present sequence represents a partial human core 2 beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide. The polypeptide can be used to treat diseases and disorders, such as cancer, cardiovascular disorders and inflammatory disorders including asthma, rheumatoid arthritis, inflammatory bowel disease, arteriosclerosis, septic shock, adult respiratory distress syndrome (ARDS) and cancer. Various platelet-mediated pathologies such as atherosclerosis and clotting can also be treated. The polypeptides of the invention are predominantly expressed in gastrointestinal tissue (stomach, colon, intestine, testis) and are elevated in cancer. Gastrointestinal disorders that may be prevented or treated include ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis and ulcerative colitis. The antibodies may be used in immunohistochemical analysis, to detect the novel polypeptide and to localize it to particular cells and tissues and to specific subcellular locations and to quantitate the level of expression.

Sequence 663 AA;

Query Match 36.3%; Score 868; DB 21; Length 663;
Best Local Similarity 43.0%; Pred. No. 1e-69;
Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;

QY 23 LMLSLKLKLVN-----RLFPQKDIYVEYSLSSTSPFVRN-RYTHVKDEVREYVNCSG 75
DB 97 lwaigcymllatvalklsfirkcdshlglesresgsqycrnllnflkpkarsincsg 156
QY 76 IV-EQEPPL-EIGKSLERIRRDIDLEDDDVAMTSDCDIYQTLRGYAQKIVSKREKSF 131
DB 157 vtrgdqeaavlqallnmlvkkrr-epftdthylsltrdcenhfaekrfiqfplskeevf 215
QY 132 PIATSLVYHKDAIMVERLHAIVNQHNICIHDKRAKAPTEFKVAMNNLAKCFSNIFTASK 191
DB 216 playsmvlnhekienferlliravypqnlcyvhvdekspetfkeavvkaaiscfpnvflask 275
QY 192 LEAVEYAHISRLQADNLCSLDLSSSIOMKYVINLGGDFPKKSNELSELKUNGAM 251
DB 276 lrvvvaswstrvgadlnmedllqssvwpkyflntcgtfdipknaemvqalkmlngns 335
QY 252 LETVPPNSKLEFRFYHHELRVRYEVYKLPFRINISKAAPPHNIQIFVGSAYVLSQAF 311
DB 336 mesevppkhketrwkyhfevr---dtlhl---tnkkkdpppylnlmtfgnayivasrdf 389
QY 312 VKYIFNNSIVQDFEAFMSKDTYSPDEHFATLIRVPGIGEI-SRSADVSDLSQKTRLVK 370
DB 390 vqhvlnknpksqqlievwkdtyspdehlwatlqratwmpsvpnphkpydisdmtslarlvk 449
QY 371 WNYEGFF-----YPSCTGSHLRSVCYGAELRMLIKDGHWFANKFDSKVDPIILKCLA 425
DB 450 wqhggedldkgapyapcsghnqratlcvygagdlmwlnqnhllanlkfddkvdnalqcle 509
QY 426 EKL 428
DB 510 eyl 512

RESULT 7
ID AAW93942 standard; Protein: 428 AA.
XX AAW93942;
AC AAW93942;
DT 28-JUN-1999 (first entry)
XX
XX Rat DH1 protein.
DE
XX
XX DH1; rat; screening; treatment; prevention; cardiomyopathy; inhibitor;
KW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;

CC proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or CC proliferative disorders. AAC9232 to AAC9240 and AAB54467 represent CC sequences used in the exemplification of the present invention.

XX Sequence 465 AA;

Query Match 36.2%; Score 866; DB 21; Length 465;
Best Local Similarity 43.0%; Pred. No. 9.3e-70;
Matches 182; Conservative 74; Mismatches 143; Indels 24; Gaps 10;

```

QY 23 LMLLS--LKLINVRLEPQK---DIYLVESLSTSPFVFN-RYHVKDEYREYVNSG 75
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40 lwalqymalatalvalklslfxlkcdedhlglesresgqyrnllynfiklpakrslnsg 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 76 IY--EOEDL--EIGKSLIRRDIDLEDDEDDVAMTSQCDIYQTLRGIAQKLVKESKF 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 100 vrrgqaeavlgallnlnlevkkr-efldthylstlceehkkaerklidgfpiskevef 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 132 PLAVSLVHKDAIMVERLHAIVNQHNYICIHDRKAPDFPKVAMNNLAKCFNSIFLASK 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 139 playamvlnhekienterilravapqqlcyvhvdekspeetkeavkaliscfpnvflask 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 132 LEAVYAHISRQADNLNLSLILKSSIQWKVYNLCQDPFLKSNFELVSELKLNAGNM 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 219 lrvvyaaswsvrqadlnmedlqgsyvpwkyflncgtdfipksnaemvqalmlnrgns 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 252 LETVAPSKLERFYHHEHLRRVPEYVKLPFRINISKAPPHNIQIVGSAVYLSQAF 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 279 mesevppkhketrwkyhvevr---dthl---cnkkdpppylnlmtgnayivasdf 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 312 VAYITNNSTVQDFEAMSKDITSPDEHFWATLIRVPGIIGEL--SSAQDVSPLOSQKTRLVK 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 333 vghvlnkpksgqllelvwtdlyspdehltwatlgtraimmpgsvpnmpkxydismtslarlvk 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 371 WMYVYGGF-----YPSCTGSHLRVCICGAELRWLIKDGHWMPKRFPSKYDPLIKCLA 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 393 wghhgdldkdgayapcsqllhgralcyvgagdlmwlmgnhlankfipkvdnaigcle 452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 426 EKL 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 453 eyl 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 9

AAM93943

AC AAM93943;

DT 28-JUN-1999 (first entry)

DE Human core 2 Gnt protein.

Screening; treatment; prevention; cardiomyopathy; inhibitor;

diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;

UDP-GlcNAc:Galbeta1-3GalNAc-alpha1; transgenic animal; germ line;

beta-1,6-N-acetylglucosaminyl-1-transferase; human; core 2 Gnt.

OS Homo sapiens.

PN CA2186987-A.

PD 02-APR-1998.

PF 02-OCT-1996; 96CA-2186987.

PR 02-OCT-1996; 96CA-2186987.

PA (MOUN) MOUNT SINAI HOSPITAL CORP.

PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;

XX WPI: 1998-399608/35.
DR N-PSDB; AAX24043.

PT Screening for substances that prevent or treat cardiomyopathy
PT associated with diabetes and hyperglycaemia - complees reacting
PT core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
PT in presence of test substance

PS Disclosure; Fig 9; 35pp; English.

CC This invention describes a method for screening for a substance that
CC may be used to prevent or treat cardiomyopathy associated with diabetes
CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
CC (UDP-GlcNAc:Galbeta1-3GalNAc-alpha1-beta1,6-N-acetylglucosaminyl-
CC transferase) with an acceptor substrate and a sugar nucleotide donor in
CC the presence of a test substance under conditions whereby the core 2
CC GlcNAc-T produces a reaction product, determining the amount of reaction
CC product, and comparing the amount of reaction product with the amount
CC obtained in the absence of the test substance, where lower amounts of
CC reaction product in the presence of the test substance indicate that the
CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
CC methods for preventing or treating cardiomyopathy associated with
CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
CC germ cells and somatic cells all contain a DNA construct introduced into
CC the animal or an ancestor of the animal at an embryonic stage, where
CC incorporation of the DNA construct into the germ line of the animal
CC causes the animal to develop cardiomyopathy similar to that associated
CC with diabetes mellitus and hyperglycaemia. This sequence represents the
CC human core 2 Gnt protein (beta-1,6-N-acetylglucosaminyltransferase)
CC which is used to describe the method of the invention.

XX Sequence 428 AA;

Query Match 36.1%; Score 863.5; DB 19; Length 428;
Best Local Similarity 43.1%; Pred. No. 1.4e-69;
Matches 187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;

```

QY 28 LKLINVRLEF--PQKDIYLV-EYSLSTSPFVNRRTYHVKDE--VRY-----EV 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 mirltllrrllfsytkyfmvlalsltfslvr---lhqkpefsvyrhlagempssdl 57
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 72 NCGIYEOEPLIEIGK-----SLFIRRDIDLEDDEDDVAMTSQCDIYQTLRGIAQKLV 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 58 nckvlgdvneiqkylalltvkfkrrp--rtvppdylnmsdcgsfiktiryivepsl 115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 KEKSPFIAYSLVHKDAIMVERLHAIVNQHNYICIHDRKAPDFPKVAMNNLAKCFNS 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 keeefplayslvvnhklemldrllralympgnfyvchvdxsedeyslaavmgjaascfen 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 186 IFIASKLEAVEYAHISRLQADNLNLSLILKSSIQWKVYNLCQDPFLKSNFELVSELK 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 176 yfvastresvvaswsvrqadlnmedlqgsyvpwkyflncgtdfipksnaemvqalmlnrgns 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 246 LGAANMLETVKPNKSLERFYHHEHLRRVPEYV--KLPIRTINISKAPPHNIQIVGSA 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 lmgennletermpshkeery-----kryevngkl-tntgtvmlpplerlpsgsa 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 304 YFVISOAFYKVIINNSIVQDFEAMSKDITSPDEHFWATLIRVPGIIGELISRSNO-DVSD 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 288 yfvastresvvaswsvrqadlnmedlqgsyvpwkyflncgtdfipksnaemvqalmlnrgns 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 363 QSKTRLVKMYVYGGF-----YPSCTGSHLRVCICGAELRWLIKDGHWMPKRFPSKYD 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 348 gavarfvkwgyfgdvskgapyppcdgvyhsvrcisfgagdlmwlmgnhlankfipkfvddvd 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 418 PILIKCIAEKLIEG 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 408 lfaigcdelnhrnk 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 10

ID	Accession	Standard	Protein	Length
AA030298	AA030298	standard	Protein	428 AA
XX	AA030298			
XX	AA030298			
XX	12-FEB-2001	(first entry)		
XX				
XX	Human heart core 2	GlcNAc-T		
XX				
XX	Human; diabetes; core 2	GlcNAc-T; cardiomyopathy; hyperglycaemia		
KW	UDP-GlcNAc:Galbeta1-3GalNAc:alpha	beta1-6		
KM	N-acetylglucosaminyltransferase			
XX				
XX	Homo sapiens			
OS				
XX	US6131578-A			
XX	17-OCT-2000			
XX				
XX	02-OCT-1997	97US-0943058		
XX				
XX	02-OCT-1996	96US-0046876		
XX				
XX	(KING/) KING G L			
XX	(NISH/) NISHIO Y			
XX	(KOYA/) KOYA D			
XX	(DENN/) DENNIS J W			
XX	(WARR/) WARREN C E			
XX				
XX	Nishio Y, Koya D, King GL, Warren CE, Dennis JW;			
XX	WPI; 2000-678642/66			
XX	N-PSDB; AAC65469			
XX				
XX	Preventing or treating cardiomyopathy associated with diabetes mellitus			
XX	and hyperglycemia, comprises administering a substance that inhibits			
XX	core 2 N-acetylglucosamine-T activity			
XX				
XX	Claim 3; Fig 9; 21pp: English			
XX				
XX	The present invention describes the human UDP-GlcNAc:Galbeta1-3GalNAc			
XX	alpha beta1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein			
XX	and coding sequences. The enzyme is associated with cardiomyopathy in			
XX	diabetes and hyperglycaemia sufferers. The invention also provides			
XX	methods for identifying inhibitors of core 2 GlcNAc-T which can be used			
XX	in treatment.			
XX				
XX	Sequence 428 AA;			
XX				
XX	Query Match	36.1%; Score 863.5; DB 21; Length 428;		
XX	Best Local Similarity	43.1%; Pred. No. 1.4e-69;		
XX	Matches 187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;			
XX				
XX	28 LKLLNVRRLF--PQKDTILV-EYLSLSPVRYNRRTYKDE--VY-----EV 71			
XX	1 MLLTLLTTRITRIFSPKLYFMVIVLSLITFSYLR--LHQKPEFVSVHLELAEPSSDI 57			
XX	72 NCGSIYEQELTEIGK-----SLTERRDIDLEDDEVDVANTSDCDIYOTLRGYAOKIVS 125			
XX	58 NCKVLVGVGQVNEQKVKLELIVTKFKRP--TWPDQYIMNTSCSSFFIKRTYIVGEPIS 115			
XX	126 KEKSPPIASLVVNHDAIMVERLLHAIYNOHNITCIYDKKADPTKVMNNILACFSN 185			
XX	116 KEEKEEPIASLVVNHKIMELIDRLITRAYMPQNFYCVHVDKSDSYLAAVMGLAASCFN 175			
XX	186 IFLASKLEAVYAHNISRLQADLNCISDLKSSIQKKYVINICGODFLKSNFETVSELKK 245			
XX	176 VFVAASRIESVYVSWETGVQADINCKDLYAMSAWKYIINCGMDFIKENLSIVLKKIKI 235			
XX	246 LQNAQMLETYKPRNSKLEKFTYHNELARKVRYUY--KRPITINSKAPRPHNIQIVGSGA 303			
XX	236 LMGEDNLETERMSHKEETW-----KRYEVLVNGKI--TGTGVMKPIECPILSGAA 287			

[illegible]


```

OS Ratus sp.
XX US6131578-A.
XX 17-OCT-2000.
XX 02-OCT-1997; 97US-0943058.
XX 02-OCT-1996; 96US-0046876.
XX (KING/) KING G L.
XX (NISH/) NISHIO Y.
XX (KOYA/) KOYA D.
XX (DENN/) DENNIS J W.
XX (WARR/) WARREN C E.
XX Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
XX WPI, 2000-678642/66.
XX N-PSDB; AAC65468.
XX Preventing or treating cardiomyopathy associated with diabetes mellitus
XX and hyperglycemia, comprises administering a substance that inhibits
XX core 2 N-acetylglucosamine-1 activity -
XX Example 1; Fig 3B; 21pp; English.
XX The present invention describes the human UDP-GlcNAc:Galbeta1-3galNAc
XX alpha beta1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
XX and coding sequences. The enzyme is associated with cardiomyopathy in
XX diabetes and hyperglycaemia sufferers. The invention also provides
XX methods for identifying inhibitors of core 2 GlcNAc-T which can be used
XX in treatment.
XX Sequence 427 AA:
SQ
Query Match 35.6%; Score 850; DB 21; Length 427;
Best Local Similarity 40.9%; Pred. No. 2, 3e-68;
Matches 118; Conservative 78; Mismatches 133; Indels 46; Gaps 12;
XX 28 LKLLVRLRLE--PQRDIYLV-EYSLSTSPFVNR-----YTHVK--DEVREVNCS 74
XX | : | | | | | : | : | | | | | : | : | : | : | : | : | : |
XX 1 mltlrrrrlfsytkyfmvvlslsllfevrlhgbkpfsvshlelbgddpsnsvnct 60
XX 75 GIVEOEPLIEGKSLERRDIDLEP-----DDVAMTSPCCDIYQTLRGYACKLVSK 126
XX | : | | | | | : | : | : | : | : | : | : | : | : | : | : |
XX 61 kvlqgdpaelqkv---kaleltvgekkrrpttrpheytmctdcasfltrkyimeplk 116
XX 127 EKSFPPLASLVYHKDAIWEHLIAIYNQHNITCIHYDRKAPDFEKVAMNNIAKCFNSI 186
XX | : | | | | | : | : | | | | | : | : | : | : | : | : | : |
XX 117 eevgfplayslvvhkldmldrllraiympnfyichvdrkaeesflaavqgiasefdm 176
XX 187 FISKIEAVEYAHISRLDNLCLSDLSKSIOMKVINLCCGDFPLKSNFELVSELKTL 246
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX 177 lvasqlesvvyaswvkadlcmkaiylmnaakylldlqmdfplknllelvtklflsf 236
XX 247 NGANMLETVK-PENSKLERETYHHELRVRYEVY--KLPIRTNISKEAPENIQIFVGS 302
XX | : | | | | | : | : | : | : | : | : | : | : | : | : | : |
XX 237 tgnslsetekmpnke-----krwkyryvvvgkl-tltygvkaaprlklrpfsgs 285
XX 303 AYFVLQSAFVKYIFNNSIYODFFAWSKDYSPDEHFWALLIRVPGIPGEISSAQ-DVSD 361
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX 286 ayfvvleeygyvlenknlqktmewagdclyspdeliwalqitpevgpslpsshkydlsd 345
XX 362 LQSKTRLVKNWNYEGF-----YPSCTGSHLRSCVCIYGAELRMLIKDGHWFANKFDSKY 416
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX 346 maavarfvkwqifegdvsgagpyppsgyhwrsvcvfygdlsswmlrkhhfankfcmv 405
XX 417 DPILIKCLAEKLEEQ 431
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX 406 dpralqcleehlrhk 420

```

```

RESULT 14
AAR71932
ID AAR71932 standard; Protein; 400 AA.
XX
XX AAR71932;
XX
XX 23-SEP-1995 (first entry)
XX
XX I-branching enzyme.
XX
XX Beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; IgM;
XX inflammation; antiinflammatory; tumor; hypersensitivity; anemia;
XX transgenic animal; I antigen.
XX
XX Homo sapiens.
XX
XX W09507020-A.
XX
XX 16-MAR-1995.
XX
XX 09-SEP-1993; 93WO-US08476.
XX
XX 09-SEP-1993; 93WO-US08476.
XX
XX (LJOL-) LA JOLLA CANCER RES FOUND.
XX
XX Bierhuizen MFA, Fukuda M;
XX
XX WPI: 1995-123182/16.
XX
XX N-PSDB; AA089201.
XX
XX Nucleic acid encoding I-branching enzyme - used to develop prods.
XX for treating e.g. inflammatory responses, tumours or
XX hypersensitivity reactions
XX
XX Disclosure; Fig 6; 56pp; English.
XX
XX Poly-A RNA isolated from human PA-1 teratocarcinoma cells was
XX reverse-transcribed and inserted into expression vector pCDNA1.
XX Plasmid DNA was used to transfect CHO-Py-leu cells. Transfected
XX cells were screened using human anti-I antigen antibodies and goat
XX anti-human IgM. Plasmid DNA was analyzed to obtain DNA (given in
XX AA089201) encoding I-branching enzyme (AAR71932).
XX
XX Sequence 400 AA:
SQ
Query Match 29.8%; Score 711; DB 16; Length 400;
Best Local Similarity 43.7%; Pred. No. 7, 6e-56;
Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;
XX 108 SDCDIYQTLRGYACKLVSEKESFPPIASLVYHKDAIWEHLIAIYNQHNITCIHYDRK 167
XX | : | | | | | : | : | : | : | : | : | : | : | : | : | : |
XX 70 ssckeylqgshlylapiaskeadlplaylmvlnhftfarlfrailmpgnlvcvnhdek 129
XX 168 APDIFKVMAMNIAKCFNSIPIASKEAVEYAHISRLDNLCLSDLSKSIOMKVINLCC 227
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX 130 atfekdavegljascfnafaskmepvvyglsrlqadlncldlsafeswkyvntc 189
XX 228 GDDPLKSNFELVSELKLNANMLETVKPPNSKLERFTY-HHHLRVRPFEYKLPRTN 286
XX | : | | | | | : | : | : | : | : | : | : | : | : | : | : |
XX 190 qgdfplknkaiyqlykqkgnltpgvlppahalgtrkyvhghlqkelsyv---lrtl 246
XX 287 ISKEAPENIQIFVGSAYFVLQSAFVKYIFNNSIYODFFAWSKDYSPDEHFWALLIRVP 346
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX 247 alkppppnlllygsayvalsrfeanfvlhdpavdllyskdtkfspdelfwlrnirp 306
XX 347 GIPGEISSAQDVSDLSKTRLVKNWNYEGF--YPSCTGSHLRSCVCIYGAELRMLINDGH 406
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX 307 gvpsmpnas-----vgnlaikwsdmed-rhgccghnyvhgiclyngldikwlvnsps 360
XX 407 WFANKFDSKVPDILIKCLAEKLEEQ 433

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 08:22:05 ; Search time 91.85 seconds

(without alignments)
3358.335 Million cell updates/sec

Title: US-09-645-192-1

Perfect score: 1362

Sequence: 1 atgaagatattcaatgctta.....atctcaactacacatcatga 1362

IDENTITY_NUC

Scoring table: Gapop 10.0 , Gapext 1.0

Searched: 351203 segs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2-6/ptodata/1/ina/5A-COMB.seq.*
2: /cgn2-6/ptodata/1/ina/5B-COMB.seq.*
3: /cgn2-6/ptodata/1/ina/5A-COMB.seq.*
4: /cgn2-6/ptodata/1/ina/5B-COMB.seq.*
5: /cgn2-6/ptodata/1/ina/PTCUS-COMB.seq.*
6: /cgn2-6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	191.8	14.1	2128	3	US-09-233-506-1
2	185.2	13.6	1807	1	US-08-118-906-13
3	185.2	13.6	1807	1	US-08-486-196-13
4	185.2	13.6	1807	1	US-08-486-135-13
5	185.2	13.6	1807	2	US-08-474-065-13
6	172	12.6	2105	1	US-07-955-041-3
7	172	12.6	2105	1	US-08-227-455-3
8	172	12.6	2105	1	US-08-472-482-3
9	172	12.6	2105	1	US-08-487-069-3
10	160.8	11.8	2102	3	US-09-063-237-3
11	121.2	8.9	378	1	US-08-118-906-1
12	121.2	8.9	378	1	US-08-486-196-1
13	121.2	8.9	378	1	US-08-486-135-1
14	121.2	8.9	378	2	US-08-474-065-1
15	92.2	6.8	378	1	US-08-118-906-3
16	92.2	6.8	378	1	US-08-486-196-3
17	92.2	6.8	378	1	US-08-486-135-3
18	92.2	6.8	378	2	US-08-474-065-3
19	53.4	3.9	192	3	US-09-233-506-9
20	42	3.1	99	1	US-08-118-906-5
21	42	3.1	99	1	US-08-486-196-5
22	42	3.1	99	1	US-08-486-135-5
23	42	3.1	99	2	US-08-474-065-5
24	37.2	2.7	7218	1	US-08-232-463-14
25	35.2	2.6	99	1	US-08-118-906-7
26	35.2	2.6	99	1	US-08-486-196-7
27	35.2	2.6	99	1	US-08-486-135-7

28	35.2	2.6	99	2	US-08-474-065-7	Sequence 7, Appl
29	34.2	2.5	362	3	US-08-675-566-8	Sequence 8, Appl
30	34.2	2.5	362	3	US-08-675-566-11	Sequence 11, Appl
31	34.2	2.5	388	3	US-08-675-566-12	Sequence 12, Appl
32	34.2	2.5	395	3	US-08-675-566-10	Sequence 10, Appl
33	34.2	2.5	4009	3	US-08-675-566-9	Sequence 9, Appl
34	34.2	2.5	4503	3	US-08-675-566-7	Sequence 7, Appl
35	34.2	2.5	7379	3	US-08-675-566-13	Sequence 13, Appl
36	34	2.5	1928	4	US-08-675-816-4	Sequence 4, Appl
37	33.8	2.5	5282	4	US-09-298-3678-5	Sequence 5, Appl
38	32.8	2.4	3977	1	US-07-794-400-2	Sequence 2, Appl
39	32.8	2.4	3977	1	US-07-794-400-13	Sequence 13, Appl
40	32.8	2.4	3977	1	US-08-041-648-1	Sequence 1, Appl
41	32.8	2.4	3977	1	US-08-217-529-1	Sequence 1, Appl
42	32.8	2.4	3977	1	US-08-197-470-2	Sequence 2, Appl
43	32.8	2.4	3977	1	US-08-397-470-13	Sequence 13, Appl
44	32.8	2.4	31571	1	US-08-323-4438-1	Sequence 1, Appl
45	32.8	2.4	53526	3	US-08-658-136-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-233-506-1
Sequence 1, Application US/09233506
Patent No. 6136580
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
APPLICANT: Yeh, Jium-Chern
TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms
TITLE OF INVENTION: Core 2, Core 4 and 1 Branches
FILE REFERENCE: P-LT 3415
CURRENT APPLICATION NUMBER: US/09/233,506
CURRENT FILING DATE: 1999-01-19
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2128
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (354)..(1670)
US-09-233-506-1

Query Match	14.1%	Score 191.8;	DB 3;	Length 2128;
Best Local Similarity	52.5%	Pred. No. 1.1e-44;		
Matches 517; Conservative	0;	Mismatches 432;	Indels 36;	Gaps 3;
QY 317	tggcaccatgattgtgacattatcagacgtcctaagaggtatgctcaaaagctgtctca	376		
DB 670	tcaccagacgtgtgacacattcaagcgtgaaggaagtcataacagttccacgtgagca	729		
QY 377	aggaggaagaaagcttccacatagcctatctctgtgtgtccacaagaatgacatag	436		
DB 730	aagaagaggggtggttcctcatctgcatctatgtgtgttccatggaattggaact	789		
QY 437	ttaagagcttaccatgctatatacaacagacacatatttctcaccatattagac	496		
DB 790	ttgaagagcctacgcagcgtgtgacgtccctcagacacatattctgttcacatgtagatg	849		
QY 497	gtaagagcctgataccttcaaggttgcataagacacattagctaaagtgtcttccacata	556		
DB 850	agaagtcctccagaaacttcaagagcggtcaagaacattattcttcttccccaatg	909		
QY 557	tttctgttcccaatagagcgtgtggaatgtccacatttccagactccagctg	616		
DB 910	tcttcaagcagtaagcgtgttcggtgtgttatgtctctcctgtgtccaggtgcaagctg	969		
QY 617	attaatattgttgcagccttctgaagtcctcaatccagtcagtggaatatgttatcaact	676		

```

Db 970 acctcaactgcacggaagactctccacagctcagtcagtcgctggaataacttcctgaata 1029
Qy 677 tctgtgggcaagaatttcccccgaagtcgaatttgaattggtgcagagtgtaaaaaac 736
Db 1030 catgtgggagcggaacttccctataaagacagatgacagatggtccagctctcaagaagt 1069
Qy 737 tcaatggagcaaatatgttggagcggtgtaaaccccccaacagtaattggaagaatca 796
Db 1090 tgaatggaggagatagcatgagtgagtgacccctcaagacacaagaaccccgctgga 1149
Qy 797 ctacacatcaactgaacttagagcggtgctctatgaatatgtgaagtlaccataaagaaac 856
Db 1150 aatataccttgaaggatgagagacacattacac-----ctaacca 1191
Qy 857 acatcccaaggaagacaccccccaataacatcagataattgttcagtgcttatttg 916
Db 1192 acaagaagaagatccctcccttataatttaactatgtttacagggagatgctacattg 1251
Qy 917 ttttaagtcacgatttcttaatatatttttaacaactccatgctcaagaacttttg 976
Db 1252 tgggtcccgagatttcgaccacatgttttgaagacccttaaacctcaaacacatgattg 1311
Qy 977 cctgtgctaaagacacatactctcctgtagagacatttgggtacacttgatcggtc 1036
Db 1312 aatgggtaaaagacacactataagccagatagacacactctggccaccccttcagcgtgac 1371
Qy 1037 caggaataccttgaggagatttccagatccagcccaagatg---tgtctatctgcagagta 1093
Db 1372 ggtgtagtgcctgtcctgtctgtcccaaccccaagtaagacatccacacatgacttcta 1431
Qy 1094 agactgccttgaagtgagtaactatgaaagcttttctcaccagtt----- 1143
Db 1432 tggccgggtgtgcaagtgagtgagtgatagaggagacatgaagggttcctctatg 1491
Qy 1144 ----tgtatggtatccacttgcgaagcgtgtatattatgagtgagctgcagaatgaagtc 1198
Db 1492 ctccctgccttgatccaccagcggtctatctgcttattgagtggtggagcttgaaatt 1551
Qy 1199 gacttcaaaagatgacacttggttgcataaatttcaattcctaagcttgagccctatct 1258
Db 1552 ggaatgcttcaaaacccctcctgttgccacaagaatttgaccacaaggttagatgataatg 1611
Qy 1259 tgattaatgcttgagcaagaagct 1283
Db 1612 cctctcagtgcttagaagaataact 1636

RESULT 2
US-08-118-906-13
; Sequence 13, Application US/08118906
; Patent No. 5484590
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Bierhuizen, Marti F.A.
; TITLE OF INVENTION: Expression of the Developmental I
; TITLE OF INVENTION: Actigen By a Cloned Human CDNA Encoding a Member of a
; TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118, 906
; FILING DATE: 09-SEP-1993

```

```

; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LI 9526
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 255..1454
; US-08-118-906-13

```

```

Query Match 13.6%; Score 185.2; DB 1; Length 1807;
Best Local Similarity 52.1%; Pred. No. 7.2e-43;
Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```

```

Qy 327 ttgtgacatttatcagctcctaagaggttatgctcaaaagctgtctcctaagagagaga 386
Db 467 TTGCAAGGAATACTTGACCCAGAGCCACTACACAGCCCTTTATCTAAGGAAGAGAGC 526
Qy 387 aagctcccaatagccatctcttgggtgcacaagaatgacatattggttgaagct 446
Db 527 TGACTTTCCCTGGCATATATATATGTCATCATCATCACTTTGACACCTTTGCAAGGCT 586
Qy 447 tatccatgcatatatacaacacagcaaatatttctgcatcattatgatggtgaagcacc 506
Db 587 CTTGAGGGCATTTTACATATGCCCCCAAAATATCTACTGTTCATGATGATGATGATGATG 646
Qy 507 tgatacttcaaatgtccatgacaattagctaaagtgtcttccaatatttctatgc 566
Db 647 AATGATTTTAAATATGCGGTAGAGCAACTATTAAGCTGCTTCCAAAGCCTTTTGGC 706
Qy 447 tatccatgcatatatacaacacagcaaatatttctgcatcattatgatggtgaagcacc 506
Db 587 CTTGAGGGCATTTTACATATGCCCCCAAAATATCTACTGTTCATGATGATGATGATGATG 646
Qy 507 tgatacttcaaatgtccatgacaattagctaaagtgtcttccaatatttctatgc 566
Db 647 AATGATTTTAAATATGCGGTAGAGCAACTATTAAGCTGCTTCCAAAGCCTTTTGGC 706
Qy 567 ttccaatttagagcgtgtggaatatgccacattccagactccaggtgatttaattg 626
Db 707 TTCCAAAGTGAACCCCTTGTCTATGAGGAGATCTCCAGGCTCAGGCTGACCTGAACCTG 766
Qy 627 ctgtcgacctctggaagcttcaatccagtggaatatgtatacaacttgtgtgga 686
Db 767 CATCAGAGATCTTCTGCTTCCTTCAGAGTCTCATGGAATGATGATATCAACACCTGTGGCA 826
Qy 687 agatttccctgaagcaaatatttgaattggtgctcagagttgtaaaaaactcaatggagc 746
Db 827 AGACTTCCCTGTAACCAACCAAGAAATAGTCACTACTCTAATAAGATTTAAGGTAA 886
Qy 747 aatcatgttgagagcgttgaaaccccccaacagtaaatgtgaagaatttccattacatca 806
Db 887 AAATATCAACCCAGAGGCTGCTGCCCCCAGCTCATGCAATTTGAGAGGACTAATATGCA 946
Qy 807 tgaacttagacggtgctcctatgaaatgtgaaacttaaccaataagaacaacatcccaa 866
Db 947 CCAAGAGACACTGTGGCA-----AAGAGCTTTCCTATGTATGAAGAACACAGCTTGAA 1000
Qy 867 ggaagacaccccccaataacattcagataatttggtagagtgctatttctttaaagta 926
Db 1001 ACCGCTTCCCTCATATATCATCAAAATTTACTTTGGCTCTGCTATGTGCTCTATCAAG 1060
Qy 927 agcatgtgtaaalatalatttcaacaacatccatcgttcaagacttttgcctggtctaa 986
Db 1061 AGAGTTTGCCAACTTGTGTGCTGATGACCCACAGGCGTGTGATTTGCTCCAGTGCCAA 1120
Qy 987 agacacatcctcctgtagagcactttgggtacacttgatcgggttccaggaataacc 1046
Db 1121 GGACACTTTCAGTCTGATGAGCATTTCTGGTGACACTCAATAGATGATTCAGAGTGTCC 1180

```

QY 1047 tgggagatttccagatcaagccagagatgtctatctgcagtaagtaactgcctgt 1106
 Db 1181 TGGCTCTATGCCAATGATCCTGACAG-----GAAACTCAGAGCTAT 1225
 QY 1107 caagtggaattactagaagcttttctatccagttgtaactgtacccctcgaag 1166
 Db 1226 AAAGTGAGGTGACATGGAAGACAGACGAGAGC---TGCCAGGCGCACTATGTACATG 1282
 QY 1167 cgtgtatcttctgagctgcagaattaggtgtctatccaagaatgacattgtgtgc 1226
 Db 1283 TATTGTATCTATGGAAGACGAGACTTAAAGTGTGTTAATTCACCAAGCTGTGTGC 1342
 QY 1227 taataattgattcttaaggtgagccctatctgtatgaagcttggaagaagcttga 1286
 Db 1343 TAACAAGTTTGAGCTTAATACCTACCCCTTACTGTGAAATGCTTAGAAGTGAAGCATCG 1402
 QY 1287 agaa 1290
 Db 1403 CGAA 1406

RESULT 3

US-08-486-196-13
 ; Sequence 13, Application US/08486196
 ; Patent No. 5731420

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
 APPLICANT: Bierhuizen, Marti F.A.
 TITLE OF INVENTION: Expression of the Developmental I
 TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESS: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA

ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/486,196

FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/118, 906
 FILING DATE: 09-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 9526
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1807 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA

FEATURE:
 NAME/KEY: CDS
 LOCATION: 255..1454
 US-08-486-196-13

Query Match 13.6%; Score 185.2; DB 1; Length 1807;
 Best Local Similarity 52.1%; Pred. No. 7.2e-43;

Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

QY 327 ttgagacattatcagacttaagaggtttagtccaagaactgtctccaagaagagaa 386
 Db 467 TTGCAAGGATCTCTGACCCAGAGCCACTACATCAGAGCCCTTTTCTTAAGAGAGAGC 526
 QY 387 aagctcccaatagactatcttctgtgtcccaagaatgcaattatgtgtgaagct 446
 Db 527 TGACTTTCCCTTGCCATATATTAATGTCATCCATCATCACTTTGACACCTTGCAAGGCT 586
 QY 447 tatcagcctataacagcaagaacaatttctgtatccatattcatatgatgtgaagacc 506
 Db 587 CTTAGGAGCTATTTCATGCCCCCAAAATATCTGTCGTTCATGTGATGATAAAGCAAC 646
 QY 507 tgatacctaaagtgtgcatagaacaattagctaaagtgtctccaataatttcatgc 566
 Db 647 AACTGAATTTAAGATGCGGTAGAGCACTATTAAAGTGCTTCCAAAGCCTTTCTGGC 706
 QY 567 ttccaattgagagctgtgtgaatatgcccacatttccaagctccaggtcgaattaatg 626
 Db 707 TTCCAAAGTGGAAACCCGTTGTCTATGAGGAGATCTCCAGGCTCCAGGCTGACCTGAAC 766
 QY 627 ctgtcgacctctgaagcttcaatccagttgaaatatgtatcaactgtgtgagaa 686
 Db 767 CATCAGAGATCTTCTGCTGCTTCAGAGTCTCATGAGAGTACGTATATCAGACCTGTGGCA 826
 QY 687 agatttccccctgaagcaaatlttgatgtgtcgaaggtgaaaaactcaatgagac 746
 Db 827 AGAATTTCCCTGTGAACCAACAGGAAATAGTTCATCTATCAAAAGATTTTAAAGTAA 886
 QY 747 aaatgtgtgagacgttgaaccccccaacagtaaatgtgaagaagttcaactacata 806
 Db 887 AAATATACCCCAAGGAGTGTCTGCCCCAGCTCATGCAATTTGACGAGACTTAATATGTC 946
 QY 807 tgaactgaacggtgtgcattgaatatgtgaagcttaacaataaggaacaactccaa 866
 Db 947 CCAAGAGCACCTGTGGCA-----AAGAGCTTCTCTATGTATGAAGAAACAACGCTTGA 1000
 QY 867 ggaagacccccccataacatcagatatgttgcagagcttatgttttaagta 926
 Db 1001 ACCGCCCTCCCCCATATATCTCACAATTTACTTTGGCTGTGCTATGTGGCTTATCAAG 1060
 QY 927 agaatgttaaatatatttccaacaactccatccgttcaagactttttgctgtgtcaa 986
 Db 1061 AGAGTTGGCAACTTGTGTCTGATGACCCACGCGCTGTGATTGCTCCAGTGGTCA 1120
 QY 987 agaacactactctcctgaatgaagcaatttgggtacttgatcgtgtccagaatacc 1046
 Db 1121 GGACACTTATAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1180
 QY 1047 tgggagatttccagatcaagccagagatgtctatctgcagtaagtaactgcctgt 1106
 Db 1181 TGGCTCTATGCCAATGATCCTGACAG-----GAAACTCAGAGCTAT 1225
 QY 1107 caagtggaattactagaagcttttctatccagttgtaactgtacccctcgaag 1166
 Db 1226 AAAGTGAGGTGACATGGAAGACAGACGAGAGC---TGCCAGGCGCACTATGTACATG 1282
 QY 1167 cgtgtatcttctgagctgcagaattaggtgtctatccaagaatgacattgtgtgc 1226
 Db 1283 TATTGTATCTATGGAAGACGAGACTTAAAGTGTGTTAATTCACCAAGCTGTGTGC 1342
 QY 1227 taataattgattcttaaggtgagccctatctgtatgaagcttggaagaagcttga 1286
 Db 1343 TAACAAGTTTGAGCTTAATACCTACCCCTTACTGTGAAATGCTTAGAAGTGAAGCATCG 1402
 QY 1287 agaa 1290
 Db 1403 CGAA 1406

RESULT 4

US-08-486-135-13

```

; Sequence 13, Application US/08488135
; Patent No. 5766910
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Biehnuzen, Marti F.A.
; TITLE OF INVENTION: Expression of the Developmental I
; TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
; TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: #1.0, Version #1.25
; APPLICATION NUMBER: US/08/488,135
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/118,906
; FILING DATE: 09-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9526
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 255..1454
; US-08-488-135-13

```

```

Query Match 13.6% Score 185.2; DB 1; Length 1807;
Best local similarity 52.1%; Pred. No. 7,2e-43;
Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```

```

; Sequence 13, Application US/08474065
; Patent No. 5830465
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Biehnuzen, Marti F.A.
; TITLE OF INVENTION: Expression of the Developmental I
; TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
; TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,065

```

FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/118,906
 FILING DATE: 09-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 9526
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1807 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 255..1454
 US-08-474-065-13

Query Match 13.6%; Score 185.2; DB 2; Length 1807;
 Best Local Similarity 52.1%; Pred. No. 7.2e-43;
 Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

QY 327 ttggacattatcagaactcctcaagaagttatgctcaaaagcttgctcacaaggagagaa 386
 DB 467 TTGCAAGGAATACTTGTGACCCAGACCCACTCATACACGCCCTTACTAAGAGAGAGC 526
 QY 387 aagcttcccaatagacctattcttggtgtccacaagaatgcaatttggttgaaggt 446
 DB 527 TGACTTTCCTTGCCATATATAATGNCATCCATCATCTTGACACCTTTGCAAGCT 586
 QY 447 tatccatgctataatacaaccagccaatattactgcataccatgatcgtgaagacc 506
 DB 587 CTTAGGGCTATTACATGCCCCAAATATCTACTGTGTTGATGATGATAAAGCAAC 646
 QY 507 tgatacttcaaatgtgcatgaacaattagcgaagtgtcttcccaataatttcattgc 566
 DB 647 AACTGAATTTAAAGATCGGAGAGCAACTATTAAAGCTGTTCCCAACCTTTTCTGSC 706
 QY 567 ttccaattgagagctgtggaatatgcccaattccagaactccaggtgattaaatg 626
 DB 707 TTCGAAGATGGAACCCGTTGTATGAGGGATCTCCAGGCTCCAGGCTGACCTGAAC 766
 QY 627 ctgtcggacctctgaagcttcaatccagtggaatatgttatcacttggtggca 686
 DB 767 CATCAGAGATCTTCTGCTTCCAGGCTCATGAGAACTACGTTATCAACCTGTGGCA 826
 QY 687 agatttcccccgaagcaaatgtgaattggtgctcagagttgaaaaaactcaatgagc 746
 DB 827 AGACTTCCCTGTAACCAACAGGAATAGTTCACTATCTGAAGATTTAAAGATA 886
 QY 747 aaatatgttgagagctggaacccccaacaaatgaattggaagatttcctacatca 806
 DB 887 AATATATACCCCAAGGCTGCTGCCCTCAGCTCATGCAATGAGAGGAGCTAAATATGCA 946
 QY 807 tgaacttaagcgggtgctcattgaataatgtgaagctaacataaaggagaacatccca 866
 DB 947 CCAAGAGACACTGGCA-----AAGAGCTTTCCTATGATATAGAACAAACACGTTGA 1000
 QY 867 ggaagcagcccccataaacttgaatattgttgccagtgctatttgtttaaagta 926
 DB 1001 ACCGCCCTCCCCCATATATGCAATTTACTTGGCTGTGCTATGTGCTTATCAAG 1060
 QY 927 agcattgttaataataatttcaacaactccatcgttcaagacttttggcctggtctaa 986
 DB 1061 AGAGTTTGCCAACTTGTCTGCATGACCCAGGCGCTGTGATTCTCTCAGTGTGCA 1120

QY 987 agacacactctccatgatgagcaatttgggtacacttgatcgggttccagaatacc 1046
 DB 1121 GGAACCTTTCAGTCCGATATAGCATTTTGGTGACACATCAATAGATTCCAGGTTTCC 1180
 QY 1047 tgggggaatttccagatcagccagagatgtgtctgactcgcagagtaagactgcctg 1106
 DB 1181 TGGCTCTATGCCAAATGCAATCGGACTG-----GAAACCTCAGACCTAT 1225
 QY 1107 caagtgaattcattgaaggttttctatccagtttactggttccacttgaag 1166
 DB 1226 AAGTGAGTGCATGGAACAGACAGAGAGC-----TGCCAGGCCACTATGTACATGC 1282
 QY 1167 cgtgtatttctgagcgtgcagaatlaaggtgcttataaagatggaatggttgc 1226
 DB 1283 TATTGTATCTATGGAACGAGACTTAAAGTGGCTGTTAATTACCAAGCCTGTTTC 1342
 QY 1227 taataattgattcgaagtgagacctatcttgatataatgcttgcagaaagcttga 1286
 DB 1343 TAAAGTTTGAGCTTAATATACCTACCCCTTACTGTGAAATGCTTAGAAGTGAAGCATCG 1402
 QY 1287 agaa 1290
 DB 1403 CGAA 1406

RESULT 6
 US-07-955-041-3
 Sequence 3, Application US/07955041
 Patent No. 5360733
 GENERAL INFORMATION:
 APPLICANT: FUKUDA, MINORU
 APPLICANT: BIERHOIZEN, MARTI PA
 TITLE OF INVENTION: A NOVEL BETA-6
 TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
 TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
 TITLE OF INVENTION: ENZYMAIC ACTIVITY
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: CAMPBELL AND FLORES
 STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
 CITY: SAN DIEGO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/955,041
 FILING DATE: 19921001
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: CAMPBELL, CATHRYN
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 9294
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-535-9001
 TELEFAX: 619-535-8949
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2105 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: both
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 220..1504
 FEATURE: polyA_signal
 NAME/KEY: polyA_signal
 LOCATION: 1913..1918


```

Db 549 TGTAAACCCCTTAAAGAGAGCGAGTTTCCAAATACATATTTCTAATAGTGCTCA 608
Qy 420 caaagatgaatattatgttgaaagcttatccatgtctatatcaaacagccaatatta 479
Db 609 TCACAAGATGTAATGCTTGACAGGCTCTGTAGGGCCATTAATAGCCTCAGAAATTTCTA 668
Qy 480 ctgcacatcattatgtatgaagcaccgtatataccttcaagtccatgaataattagc 539
Db 669 TTGCGTTATGTGACACAAAATCCGAGCATTTCTATTACTGACATGAGTGGCGATCGC 728
Qy 540 taagtgcttcccaatatttccatgtcttcccaattgaagctgtggaataatgcccacat 599
Db 729 TTCCGTTTATGATATGCTTTTGCCGACCGCATGAGAGCTGTGTTTATGATGCTG 788
Qy 600 ttccgaactccagctgtatataattgtcttgcgaaccttgcgaagcttcaatccagtg 659
Db 789 GAGCGGGGTGACGGCTGACCTCACTGATGAGGATCTCTATGCAATGAGTGCNAACGTG 848
Qy 660 gaaatgttatcaacttgctgttggaagatttcccttggaagcaaatlttgatgtg 719
Db 849 GAAGTACTGATTAATCTTTGTGTGATGATTTTCCATTAAACCACTAGAAATTTGT 908
Qy 720 gtcaagattgaaaaaactcaatggaagcaataatgttggaagcgggtgaaccccaacag 779
Db 909 CAGGAAGCTCAAGTTGTTAATGGGAGAAAACACCTGSAACGGAGAGATGCATCCCA 968
Qy 780 taatgtgaagatcattacatcactcaatgaactagaaggtgacctatgatatgtga 839
Db 969 TAAAGAAAGAGGTGGAAGAACGGGTATGA-----GGTCTTAAATGGAAGAGCTGAC 1019
Qy 840 gctaccaataaagcaaacatctccaaaggaagcaccocccataacattcagatatgt 899
Db 1020 -----AAACACAGGAGACTGCNAATACTTCCTCCATCGAACACCTCTCTTTC 1070
Qy 900 tggcagtgctatttgttttaagtaagccttgttataatatttccaacatccat 959
Db 1071 TGGCGTCCCTAATCTCGGCTGACGAGGAGATGTGGGATGTACTACAGAAATGAAA 1130
Qy 960 cgttaagacttttctgctgtcttaagacacatactctctgatgagcactttggc 1019
Db 1131 AATCCAAAAGTTGATGAGTGGGACAAAGACACATACGCCCTATAGTATCTCTGGCC 1190
Qy 1020 taccttgatcggttccaggaataaccctgggaat---ttccagatcagcccaagatg 1076
Db 1191 CACCATCCAAAGATTCCTGTAAGTCCGGGCTCACTCCCTCCGACCAATAGATGATCT 1250
Qy 1077 gctgatctgaagaaagactgcctgttcaagtggaataactgaagcctttt--- 1133
Db 1251 ATCTGACATGCAAGACATGTCAGAGTTGTCAAGTGGCAATGAGGTGATGTTTC 1310
Qy 1134 -----ctatcccaatttactgtatctcacccttgcgaagcgtgttatttgg 1181
Db 1311 CAAGGGTGTCTCCATCCCGCCTCGATGAGTGCATGTGGCTGCTGATGATTTTCGG 1370
Qy 1182 agctgcagaattaaagtggtcttcaagaagatgacatgtgttgcataaatttgatc 1241
Db 1371 AGCTGGAGACTTGAACGAGATGCGCCCAACACCACTTGTGTTCCCAATTAAGTTGACGT 1430
Qy 1242 taagtgagacctatcttgattaaatgcttggacaga 1277
Db 1431 GGATGTTGACCTCTTGGCCATCCAGTGTGATGAGA 1466

```

```

RESULT 8
US-08-472-482-3
; Sequence 3 Application US/08472482
; Patent No. 5658778
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETA1-6
; TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,

```

```

; TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMATIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,482
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERYN
; REGISTRATION NUMBER: 31,815
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 220..1504
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1913..1918
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 248..314
; OTHER INFORMATION: /standard.name-
; OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-472-482-3
;
; Query Match
; Best Local Similarity 12.6%; Score 172; DB 1; Length 2105;
; Matches 515; Conservative 0; Mismatches 445; Indels 36; Gaps 4;
;
; Qy 300 tggatgtgtgttggaatgacagatgtgtacatttcaagctctaagaggtatgc 359
; Db 489 TGACGCTATATAACATGACAGTGCATGCTCTTCTTTCATCAAGAGACGCAATATAT 548
; Qy 360 tcaaaagctgtctcaaaagaggaagaagcttcccaatagctattcttgggtgcca 419
; Db 549 TGTAGAACCCCTTAGTAAGAAAGAGCGGAGATTTCGAATAGCATATTCATAGTGCTTCA 608
; Qy 420 caaagatgaatattatgttgaaagcttatccatgtctatatcaaacagccaatatta 479
; Db 609 TCACAAGATGTAATGCTTGACAGGCTCTGTAGGGCCATTAATAGCCTCAGAAATTTCTA 668
; Qy 480 ctgcacatcattatgtatgaagcaccgtatataccttcaagtccatgaataattagc 539
; Db 669 TTGCGTTATGTGACACAAAATCCGAGCATTTCTATTACTGACATGAGTGGCGATCGC 728
; Qy 540 taagtgcttcccaatatttccatgtcttcccaattgaagctgtggaataatgcccacat 599
; Db 729 TTCCGTTTATGATATGCTTTTGCCGACCGCATGAGAGCTGTGTTTATGATGCTG 788

```



```

QY 600 ttccagactccagagctgttaattgcttgcgaccttcgtgaagcttccatccagtg 659
DB 789 GAGCGGGGTTCAGGCTGACCTCAACTGATGATGATGATGATGATGATGATGATGAT 848
QY 660 gaataatgttccaaactgtgtggaagaatttccctgaagctcaatttgaattgt 719
DB 849 GAAGTACTGATTAATCTTTTGGTATGATGATGATGATGATGATGATGATGATGAT 908
QY 720 gtcaagattgaaaaaactcaatggaagaatactgtggaagcgggtgaaccccaacag 779
DB 909 CAGGAAGCTCAAGTGTATAAGGAGAAACCAACTGGAACGGAGAGATGATGATGATGAT 968
QY 780 taattggaagatcactcactcactcactcactcactcactcactcactcactcact 839
DB 969 TAAAGAAAGAGGTGAAGGAGGATGATGATGATGATGATGATGATGATGATGATGAT 1019
QY 840 gctacaaataaggaacaacatctccaaaggaagcaaccccccatacattcagatattgt 899
DB 1020 -----AAACACAGGAGCTGCTCAAAATGCTTCTCCTCACTCAAAACCTCTCTTTTC 1070
QY 900 tggcagctctatttctgttttaagtcagacattgttaataatatttcaaacctccat 959
DB 1071 TGGCAGTGTCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1130
QY 960 cgtcaagacttttctgctgtgctcaagaacacactcctcctgataagcactttgagc 1019
DB 1131 AATCCAAAAGTTGATGAGATGAGGACACACATACAGCCCTGATGATGATGATGATGAT 1190
QY 1020 taactgtattgggttccaggaatacctgtggagat---ttccagatagccagagatgt 1076
DB 1191 CACCAATCCAAAAGATTCCTGAAGTCCCGGCTCCTCCTCCTCCTCCTCCTCCTCCT 1250
QY 1077 gctcatctgcaggaataagactcgcctgtcgaagtgaattactatgaagcctttt--- 1133
DB 1251 ATCTGACATGCAAGACAGATGCGCAGGTTTCTCAAGTGCAGTACTTGAAGGATGATGAT 1310
QY 1134 -----ctaccagattgtactgtatctacacttgaagcgtgttattatg 1181
DB 1311 CAGAGGTGCTCCCTACCGCCCTCGATGAGATGATGATGATGATGATGATGATGATGAT 1370
QY 1182 agctcagaatttaagtgctgtatcaaaagtgcacattggttggtaataattgattc 1241
DB 1371 AGCTGTGACTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1430
QY 1242 taagtgagccctatctgttaattgcttgcaga 1277
DB 1431 GGATGTGACCTCTTGGCATCCAGTGTGATGATGATGATGATGATGATGATGATGAT 1466

```

RESULT 9
 US-08-487-069-3
 Sequence 3, Application US/08487069
 Patent No. 5684134

GENERAL INFORMATION:
 APPLICANT: FUKUDA, MINORU
 APPLICANT: BIERHUIZEN, MARTI FA
 TITLE OF INVENTION: A NOVEL BETAI-6
 TITLE OF INVENTION: N-ACETYLDILUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
 TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
 TITLE OF INVENTION: ENZYMACTIC ACTIVITY
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESSES:
 ADDRESS: CAMPBELL AND FLORES
 STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
 CITY: SAN DIEGO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,069
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/955,041
FILING DATE: 01-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHERYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9294
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2105 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 220..1504
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1913..1918
FEATURE:
NAME/KEY: misc_signal
LOCATION: 248..314
OTHER INFORMATION:
OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-487-069-3

Query Match      12.6%; Score 172; DB 1; Length 2105;
Best Local Similarity 51.7%; Pred. No. 4,1e-39;
Matches 515; Conservative 0; Mismatches 445; Indels 36; Gaps 4;

```

```

Db 969 TAAAGAGAAAGGTGGAAGGAGGTATGA-----GGTCGTTATGGAAGGCTGAC 1019
QY 840 gctacaaataaagaaacacatctccaagaagacccccataacatcagatatgt 899
Db 1020 -----AAACACAGGAGACTGTCAAAATGCTTCTCCATCGAAACCTCTTTTC 1070
QY 900 tggcagtgctatttggctttaagtcacgaatgtttaatatatttcaacaactcoat 959
Db 1071 TGGCAGTCCCTACTTCGTGTCAGTGAAGATATGTGGGTGATGACTACAGAAATGAAA 1130
QY 960 cgttaagaacttttggctgtcttaagaacacatctctctatgagacactttgggc 1019
Db 1131 AATCCAAAAGTTGATGAGATGGCGCACAGACACATACAGCCTTATGACTTCTTGCGC 1190
QY 1020 taccttgatcgggtctccaggaataacctggggagat---ttccagatcagccagatgt 1076
Db 1191 CACCATCCAAAGGATTTCTGTGAAGTCCCGGCTCCTCCTCCAGCATAAGTATGATCT 1250
QY 1077 gtctgactcgaagtaagactcgtcgtctgaagtgaatgaatgaagctttt--- 1133
Db 1251 ATCTGACATGCAAGACAGTGGCCAGGTTGTCAAGTGGAGTACTTGTGAGGTTGATGTTTC 1310
QY 1134 -----ctatcccaagtgttactgagatctcaccttcgaagcgtgtgtattatg 1181
Db 1311 CAAGGTCCTCCCTACCCGCTCGATGAGATGCCATGTGCGTCAAGTGCATTTTCGG 1370
QY 1182 agctgcagaattaggtgtgtctatcaagaatgagacatgtgtgttctataaattgattc 1241
Db 1371 AGCTGGTACTTGAACTGAGTGTGCGCAACACCACTGTTTCCCAATAAGTTTGACGT 1430
QY 1242 taagttgagccctactctgtattaatgcttggcaga 1277
Db 1431 GGATGTTGACCTCTTGCCATCCAGTGTGTGAGTA 1466

```

RESULT 10

US-09-063-237-3
Sequence 3, Application US/09063237
Patent No. 6124267

GENERAL INFORMATION:

APPLICANT: McEwen, Rodger P.
APPLICANT: Cummings, Richard D.
TITLE OF INVENTION: O-Glycan Inhibitors of Selectin Mediated
TITLE OF INVENTION: Inflammation Derived from PSL-1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 West Peachtree
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30306-3450

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/063,237
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/649,802
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRFL10CIP7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)873-8794
TELEFAX: (404)873-8795

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-063-237-3

```

Query Match 11.8%; Score 160.8; DB 3; Length 2102;
Best Local Similarity 51.0%; Pred. No. 5.9e-36;
Matches 508; Conservative 0; Mismatches 452; Indels 36; Gaps 4;

```

QY 300 tgaatgltgttggcaatgacacagatgtatgtgacattatcaagacttaagaggttatgc 359
Db 488 TGACGACTATATATACATGACGAGTGTCTTTTCATCAAGACGCAATATAT 547
QY 360 tcaaaagcttcttcaaaaggaagaaagcttcccaatagcctattcttgggtcca 419
Db 548 TGTAGAACCCTTAGTAAAGAGAGCGGAGTTTCCAAATAGCATATTTATGTGTTCA 607
QY 420 caaagatgcaatattgttgaaggtcttccatgctatatacaaccagacaatatga 479
Db 608 TCAAGATATGAATGCTTTACAGCGCTGCGAAGGCCATCTATATCCTCAGATTTCTA 667
QY 480 ctgcacacattatgactgaagacaccgtatcacttcaaaagtgtccatgaacaatttagc 539
Db 668 TTCCGTTCAATGTGACACAAAATCCGAGATTCTCATTTTAGCTGCGATGATGGCATCG 727
QY 540 taagtcttcccaatttcatcttcccaattagaagcgtgtgaatagccacaact 599
Db 728 TTCTGTTTAGTATGTCCTTTGTGCGCACCGCATTTGAGAGTGTGTTTATGCATCG 787
QY 600 ttccagaactccaggtcgaatttaattgtctgtcgaaccttgaagcttccaactcagtg 659
Db 788 GAGCCGGGTTCAAGCTGACCTCACTGATGAAGATCTCTATGCAATGAGTGAACATG 847
QY 660 gaatatgtatacaacttgtgtgtgagcaaatgttcccccgaagcgaatttgaattgt 719
Db 848 GAGTACTTGATTAATCTTTTGTGTGATGATTTTCCATTAACCACTGATAATTT 907
QY 720 gtcagaattgaaataaactcaatgagcaaatatgttgaagcgttgaaccccccaaacag 779
Db 908 CACCAAGCTCAAGTTGTATGAGGAGCAAAACCACTGGAAGAGAGATGCATCCCA 967
QY 780 taattggaagattcacttaccatcatgaacttagaagcgtgtgcttataatagtga 839
Db 968 TAAAGAACAAAGGTGAAGAGCCCTATGA-----GGTCGTTAATCGAAAGCTG-- 1016
QY 840 gctacaaataaagaaacacatctccaagaagagaccccccaataacatcagatatgt 899
Db 1017 -----ACAAACACAGGAGCTGTCAAAATGCTTCTCCACATCGAAACCTCTTTTC 1069
QY 900 tggcagtgcttatttggctttaagtcacgaatgtttaatatatttcaacaactcoat 959
Db 1070 TGGCACTGCTACTTCGTGCTCAGTAGGACATATGAGGAGTGTGACTCAAAATGAAA 1129
QY 960 cgttaagaacttttggctgtcttaagaacacatactctctgtagaacacttttgggc 1019
Db 1130 AATCCAAAAGTTGATGAGATGGCGCACAGACACATACAGCCTTATGACTTCTTGCGC 1189
QY 1020 taccttgatcgggtctccaggaatcaccgtggagat---ttccagatcagccagatgt 1076
Db 1190 CACCATCCAAAGGATTTCTGTGAAGTCCCGGCTCCTCCTCCAGCATAAGTATGATCT 1249
QY 1077 gtctgactcgaagtaagactcgtcgtctgaagtgaatgaatgaagcgtttt--- 1133
Db 1250 ATCTGACATGCAAGACAGTGGCCAGGTTGTCAAGTGGAGTACTTGTGAGGTTGATGTTTC 1309
QY 1134 -----ctatcccaagtgttactgagatctcaccttgaagcgtgtgtattatg 1181
Db 1310 CAAGGTCCTCCCTACCCGCTCGATGAGATGCCATGTGCGCTCAGTGTGATTTTCGG 1369

```



```

Db 13 CCTTATCTAAGGAAGAAGCTGACTTCCCTGGCATATATATGTCATCATCATCAC 72
QY 427 gcaattatggttgaaggcttatccatgctatatacaaccgagcaattatgctac 486
Db 73 TTGACACCTTTGCAAGGCTCTTCAAGGCTATTACATGCCCAAAATATCTGTGTT 132
QY 487 cattatgctgaaggcctgataccttcaaaagtgtccatgaaacaaattagctaa 546
Db 133 CATGTGATGAAAAAGCAACAAGTAATTAAGATGGGGTAGAGCAACTATTAGCTGC 192
QY 547 ttcccaaatatttcattgctcttccaaatlaaggctgtggaattgtcccaattcca 606
Db 193 TTCCCAAAAGCTTTTCTGCTTCCCAAGATGGAACCCGTGTATGAGAGGATCCAGG 252
QY 607 ctccagctgatttaattgctgtcgaacctctcgaagcttcaatcccaagtggaat 666
Db 253 CTCACAGGCTGACCTGAAGTGCATCAGAGATCTTCTGCTTCAGAGTCTCATGCAATAC 312
QY 667 gtatacaactgtgtgggaagatttccctcgaagccaatttgaattgtgtcagag 726
Db 313 GTATCAACACCTGTGGCAAGACTTCCCTCGAAAAACAAGAAATAGTTCAGTAT 372
QY 727 ttgaaa 732
Db 373 CTGAAA 378

```

```

RESULT 13
US-08-488-135-1
Sequence 1, Application US/08488135
Patent No. 5766910
GENERAL INFORMATION:

```

```

APPLICANT: Fukuda, Minoru
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
NUMBER OF SEQUENCES: 14
NUMBER OF SEQUENCES: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,135
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

```

```

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS

```

```

LOCATION: 1..378
US-08-488-135-1

```

```

Query Match 8.9%; Score 121.2; DB 1; Length 378;
Best Local Similarity 58.2%; Pred. No. 4,2e-25;
Matches 213; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```

```

QY 367 ctgtgtcgaaggaggaagaaagcttcccaatagcctattcttggttgccaaagat 426
Db 13 CCTTATCTAAGGAAGAAGCTGACTTCCCTGGCATATATATGTCATCATCATCAC 72
QY 427 gcaattatggttgaaggcttatccatgctatatacaaccgagcaattatgctac 486
Db 73 TTGACACCTTTGCAAGGCTCTTCAAGGCTATTACATGCCCAAAATATCTGTGTT 132
QY 487 cattatgctgaaggcctgataccttcaaaagtgtccatgaaacaaattagctaa 546
Db 133 CATGTGATGAAAAAGCAACAAGTAATTAAGATGGGGTAGAGCAACTATTAGCTGC 192
QY 547 ttcccaaatatttcattgctcttccaaatlaaggctgtggaattgtcccaattcca 606
Db 193 TTCCCAAAAGCTTTTCTGCTTCCCAAGATGGAACCCGTGTATGAGAGGATCCAGG 252
QY 607 ctccagctgatttaattgctgtcgaacctctcgaagcttcaatcccaagtggaat 666
Db 253 CTCACAGGCTGACCTGAAGTGCATCAGAGATCTTCTGCTTCAGAGTCTCATGCAATAC 312
QY 667 gtatacaactgtgtgggaagatttccctcgaagccaatttgaattgtgtcagag 726
Db 313 GTATCAACACCTGTGGCAAGACTTCCCTCGAAAAACAAGAAATAGTTCAGTAT 372
QY 727 ttgaaa 732
Db 373 CTGAAA 378

```

```

RESULT 14
US-08-474-065-1
Sequence 1, Application US/08474065
Patent No. 5830465
GENERAL INFORMATION:

```

```

APPLICANT: Fukuda, Minoru
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
NUMBER OF SEQUENCES: 14
NUMBER OF SEQUENCES: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,065
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001

```

TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 378 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..378
 US-08-474-065-1

Query Match 8.9%; Score 121.2; DB 2; Length 378;
 Best Local Similarity 58.2%; Pred. No. 4.2e-25;
 Matches 213; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Db 367 ctgtctcaagaaggagaaagcttcccaatagacctcttctgtgtgtcccaagaat 426
 13 CTTTATCTAAGAGAAGAGCTGACTTCCCTTGCCATATATATGGTCATCCATCATCAC 72
 QY 427 gcaatttggttgaaggcttatccatgcatatatacaaccagacacatatttctgcatc 486
 Db 73 TTTGACACCTTTGCCAAGGCTCTTCAAGGCTATTTCATGCCCCCAAAATATCTACTGT 132
 QY 487 cattaatgctgaagagacacctgataccttcaaatgtgcatgaacaattagctaaagtgc 546
 133 CATGTGATGAAGAAAGCAACAACTGAATTAAAGATGCGTAGAGCAACTATTAAAGCTGC 192
 Db 547 ttcccaaatatttcatgcttcccaatagaagctgtgtaatatagcccaatttcaga 606
 193 TTCCCAACGCTTTCTGCTCCCAAGATGAGAACCCGTTGCTATGAGAGGATCTCCAGG 252
 QY 607 ctccaggtctgaatgaatgctgttcgagaccttctgaagcttcaatccatcagttgaaat 666
 253 CTCAGGCTGACGTAAGTGCATGAGATCTTCTGCTCGAGGCTCATGGAAGTAC 312
 QY 667 gttatacaactgtgtggaagatatttccctgaagatcaaatattgattgtgtcagag 726
 Db 313 GTATCAACACCTGTGGCAAGACTTCCCTGAAACCAACAGGAATATGTTCACTAT 372
 QY 727 ttgaag 732
 Db 373 CTGAAA 378

RESULT 15
 US-08-118-906-3

Sequence 3, Application US/08118906
 Patent No. 5484590

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru

APPLICANT: Blerhuizen, Marti F.A.

TITLE OF INVENTION: Expression of the Developmental I

TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a

TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/118-906

FILING DATE: 09-SEP-1993

CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 9526
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 378 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..378
 US-08-118-906-3

Query Match 6.8%; Score 92.2; DB 1; Length 378;
 Best Local Similarity 52.8%; Pred. No. 6.3e-17;
 Matches 199; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 355 tatgtccaaagctgttcccaaggagaaagcttcccaatagacctcttctgtgtgtcccaagaat 414
 1 TATATTGTGAACCCCTTAGTAAGAAGAGCGGAGTTCCAAATGCAATTTCTATAGTG 60
 Db 415 gtccacaagaatgcaattatgttgaagagcttatccatgcatatacaaccagacaaat 474
 61 GTTCATCAACAAGTTGAATGCTTGACAGCTGCGAGGCGCATGTATATGCTCAGAAAT 120
 QY 475 attactgcatcattatgatcatgtaagagacctgataaccttaagttgcatgaacat 534
 121 TTCTATTGGTTCTATGTGACACCAAAATCCGAGATTCTTATTAGCTGACAGTGGGC 180
 Db 535 tttagcaagtgcttcccaatatttcatgcttcccaattagaagctgtggaatatgcc 594
 181 ATGCGTTCCTGTTTAGTAAATGCTTTGGCCAGCCGATGTGAGAGTGTGTTATGCA 240
 QY 595 cacatttceagactccaaggtgatttaattgctgtggaacctctgaagcttcaatc 654
 241 TCGTGAAGCCGGGTTCAGGCTGACCTCAACTGATGAAGATCTATGCAATGAGTGCA 300
 QY 655 cagtggaatatgtatacaactgtgtggaagatttccctgaagtcgaatttgaa 714
 Db 301 AACTGGAAGTACTGATAAATCTTTGTGTATGATTTTCCCATTAACCAACCTAGAA 360
 QY 715 ttgtgtcagagttgaa 731
 Db 361 ATTGTCAAGAGCTCAA 377

Search completed: May 2, 2002, 10:30:28
 Job time: 7703 sec

Fri May 3 10:57:52 2002

us-09-645-192-1.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 08:20:30 ; Search time 1980.83 seconds
(without alignments)
11343.305 Million cell updates/sec

Title: US-09-645-192-1
Perfect score: 1362
Sequence: 1 atgaagatatacaatgta.....attccactacacatcatga 1362

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8246589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_da.*
2: gb_hcga.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pal.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_cm.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vl.*
30: em_hcgo_hum.*
31: em_hcgo_inv.*
32: em_hcgo_rod.*
33: em_hcgo_hum.*
34: em_hcgo_inv.*
35: em_hcgo_rod.*
36: em_hcgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1362	100.0	1362	6 AX087935	AX087935 Sequence
2	1362	100.0	3435	9 AF132035	AF132035 Homo sapi
3	1358.8	99.8	184590	2 AC093259	AC093259 Homo sapi
4	207.2	15.2	3508	10 AB037596	AB037596 Homo sapi
5	194.8	14.3	108873	14 AF138573	AF138573 Bovine he
6	194.4	14.3	1206	10 AB037597	AB037597 Mus muscu
7	191.8	14.1	1317	6 AX087948	AX087948 Sequence
8	191.8	14.1	1317	9 AF038650	AF038650 Homo sapi
9	191.8	14.1	2128	6 AR136078	AR136078 Sequence
10	191.8	14.1	2128	6 AF102542	AF102542 Homo sapi
11	191.8	14.1	2216	6 AX045249	AX045249 Homo sapi
12	191.8	14.1	181983	33 AC022480	AC022480 Homo sapi
13	191.8	14.1	187760	2 AC092755	AC092755 Homo sapi
14	191.6	14.1	2017	14 AF231105	AF231105 Bovine he
15	191.2	14.0	2393	10 MM068182	MM068182 Mus musculu
16	187.4	13.8	87636	9 AL139039	AL139039 Homo sapi
17	187.4	13.8	176544	2 AL138877	AL138877 Homo sapi
18	187.4	13.8	196788	2 AC021626	AC021626 Homo sapi
19	185.2	13.6	1203	6 AX087950	AX087950 Sequence
20	185.2	13.6	1807	6 I15889	I15889 Sequence 13
21	185.2	13.6	1807	6 I15029	I15029 Sequence 13
22	185.2	13.6	1807	9 HSNBAGST	HSNBAGST Sequence 13
23	185.2	13.6	1807	9 HUMNAGAT	HUMNAGAT Sequence 13
24	184.2	13.5	157102	2 AC076970	AC076970 Homo sapi
25	183.6	13.5	155359	9 AL358777	AL358777 Homo sapi
26	183.6	13.5	196788	2 AC021626	AC021626 Homo sapi
27	181.2	13.3	252504	2 AC073815	AC073815 Mus muscu
28	180.8	13.3	2217	6 AX045251	AX045251 Sequence
29	180.8	13.3	2217	6 AX045253	AX045253 Sequence
30	179.6	13.2	2033	10 MM019265	MM019265 Mus musculu
31	175.2	12.9	159930	2 AL138757	AL138757 Homo sapi
32	175.2	12.9	164041	2 AL590664	AL590664 Homo sapi
33	175.2	12.9	183858	9 AL161626	AL161626 Homo sapi
34	174.4	12.8	5010	10 S79797	S79797 enzymatic 9
35	174	12.8	2060	9 HSGNT1	HSGNT1 Human Beta
36	173.2	12.7	2023	10 D87333	D87333 Mus musculu
37	173.2	12.7	2047	10 D87332	D87332 Mus musculu
38	172.8	12.7	4179	14 BHVAREBF	BHVAREBF Bovine he
39	172	12.6	1287	6 AX087946	AX087946 Sequence
40	172	12.6	2105	6 I41262	I41262 Sequence 3
41	172	12.6	2105	6 I61440	I61440 Sequence 3
42	172	12.6	2105	6 I72548	I72548 Sequence 3
43	172	12.6	2110	9 HUMGOGG	HUMGOGG Human Beta
44	172	12.6	2204	9 HUMC2CNT	HUMC2CNT Homo sapien
45	164.8	12.1	2435	4 BTU41320	BTU41320 Bos taurus

ALIGNMENTS

RESULT 1
AX087935 LOCUS AX087935
DEFINITION Sequence 1 from Patent WO0114535.
ACCESSION AX087935
VERSION AX087935.1 GI:13396913
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1362)
AUTHORS Schwientek,T. and Clausen,H.
TITLE UDP-n-acetylglucosamine: galactose--g(b)1.3-n
-acetylglucosamine--g(a)-r/(glucan to galnac) -g(b)1.6-n
Patent: WO 0114535-A 1 01-MAR-2001;
Schwientek, Tilo (DK) ; Clausen, Henrik (DK)

FEATURES
SOURCE Location/Qualifiers
1..1362

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
1.1362
/note="The gene of C2GNT3"
/codon_start=1
/protein_id="CAC34674.1"
/db_xref="GI:13396914"
/translation="MKFRKCYFKHTLQOKYFLFILMTLSILKLNVRLPQKDI
LVERSTSPFVNRNTHVDEVRVENCSCGITEQEPLEIGKLELRNDIILED
VVAITSDCLYQTLRGRVAQKLVSKESKSPFLASLVHDAVLERLILAIYQHNII
CIHYDRAPDPTFFVANNMLAKCSNIFIAKLEAVEYAHLSRLQADNLCLDLKSI
QWYVIMLCODEPPLKSNFELVSELKLNANMLETVKPNSELEFTHHELRVY
EYVLPRTNISKSEAPPHNIQIFVGSAYFVLSQAPVKYIENNSIVODPFMSDTPY
DEHEMATLIRVPGIPGEISRSADVSQDKTRVKNWYEGEFYSQSGSHRSYCI
YGAELFMILKDGWPFANKRDSKVDPIILNCLKLELEQQRDMITLPSKELPMRNL
TTS"
BASE COUNT      410 a      260 c      279 g      413 t
ORIGIN
Query Match      100.0%; Score 1362; DB 6; Length 1362;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaagataatcaaatgttattttaacataccctacagcaaaagtttcatcctgtt 60
Db 1 ATGAAGATATCAAAATGTTATTTAAACATACCTACAGCAGAAAGTTTCATCCTGTT 60
QY 61 ttaaccatagtgctctcttctttaaagctttaaagtgagagacgtcttcggcaa 120
Db 61 TTAACCTATGAGCTCTCTCTTTGTTAAAGCTTCTAAATGAGACGACCTTTCCGAA 120
QY 121 aaagacattacttggttgagtaactccctaagtaactcgccttcttgtaagaacagatac 180
Db 121 AAAGACATTACTTGTTGAGTACTCCTAAGTACTCGCTTTGTTAGAAACAGATAC 180
QY 181 actcatgttaagatgaagatcagatgaatgaatgaatgttcgggtatctatagaacagag 240
Db 181 ACTCATGTTAAGATGAAGTCAAGGTATGAAGTGTGGGATCTTATGAACAGAGAG 240
QY 241 ccttggaaatgtgaaagagctcgtgaaataagaaagagacatcatgacttgagagat 300
Db 241 CCTTGGAAATGTGAAAGAGCTGTGAAATTAAGAAAGAGACATCATGACTTGGAGAT 300
QY 301 gatgatgttggcgaatgaacagatgatgtgacattatctcagactcgaagagttaagct 360
Db 301 GATGATGTGTCGCAATGACAGTATGTCATTAATACAGACTTAAGAGGTTATGCT 360
QY 361 caaaagcttgcctcaaaagagagaaagctcccaatagacctattcttggctgcac 420
Db 361 CAAAAGCTTGCTCAAAAGAGAGAAAGGCTTCCCAATAGCCTATCTTGGTGTGCAC 420
QY 421 aaagatgcaattatgttgaaagccttaccatctatatacaaacagacacataattac 480
Db 421 AAAGATGCAATTATGTTGAAAGGCTTATCCATGCTATATACAAACACCAATAATTAC 480
QY 481 tgcacccattatgacgtcaagcaccatgataccctcaaatgagttgcatgaacattagct 540
Db 481 TGCACCATTAATGATCGTAAAGCACCATGATACCTTCAAAAGTTGCATGAACATTTAGCT 540
QY 541 aagtccttcctcaaatatttcatgtcttccaaatagaagcgtgagaaatgaccacatt 600
Db 541 AAGTCCTCTCCAAATATTTTCATGTCTTCCAAATAGAGGCTGAGAAATAGCCACACTT 600
QY 601 tccagactcgaagctgatttaattgcttgcgaccttcgaaagcttccatccagtgag 660
Db 601 TCCAGACTCCAGCTGATTAATGCTGTGCGACCTTCGAACTCTTCATATCCAGAGG 660
QY 661 aaatatgtatcaacttggtgagcaagatttccctcgaagtcgaatcttgaattgagtg 720
Db 661 AAATATGTTATCAACTTGTGTGGGCAAGATTTTCCCTCGAAGTCAAAATTTGAAATGGTG 720
QY 721 tcagagttgaaaaaactcaatgagcaaatatgttgagacagtgaaacccccaacagct 780

```

```

Db 721 TCAGAGTTTAAAAAAGCTCATATGAGCAAAATATGTTGAGACGCTGAACCCCAACAGT 780
QY 781 aaattgaaagattaccatccatcatgaaactagaacgggtgocctatgaatcatgtgaag 840
Db 781 AAATTGAAAGATTCACTTACCATCATATGAAGCACTTGAAGCGGCTGCTTATGAATATGGAAG 840
QY 841 ctcccaataagagcaaacatctcccaaggaagacccccccataacatcagaattatgtt 900
Db 841 CTACCAATAAGGCAACAACATCTCAAGGAAGACACCCCAATCAATTCATGATATTGTT 900
QY 901 ggcagtgcttatttctttaaagcaagcattgttaatatattttcaaacactccatc 960
Db 901 GCGAGTGCTATTATTTGTTTAAGTCAAGATTTGTTAAATATATTTTCAACAACATCCATC 960
QY 961 gtccaagacttttctcctggtctaaagacaatctcctcctgatgagacatttggact 1020
Db 961 GTTCAAGACTTTTTCCTGCTCTAAAGACATCTCTCTGATGAGCACTTTGGGCT 1020
QY 1021 acctgattcgggttcacgaataacctgggagagattccagatcagccagatgttgtct 1080
Db 1021 ACCTTGATTCGGGTTTCAGGAATACCTGGGAGATTTCCAGATCAACCCAGATGTGTCT 1080
QY 1081 gatctcagagtaagacatcgctctgtcaagtgaattactatgaagcgttttctatccc 1140
Db 1081 GATCTCAGAGTAAGACTCGCCTTGTCAAGTGAATTACTATGAAGGCTTTTCTATCCC 1140
QY 1141 agttgactgagatcacccttcgaagcgtgtgattttagagagctcgcgaataaagtgag 1200
Db 1141 AGTTGACTGAGATCTACCTTCGAAGCGTGTGATTATTAAGAGCTGCACAAATTAAGGAG 1200
QY 1201 ctatacaaatgtgacattgtgtcttaataattgattctcaagtgagaccatactgtg 1260
Db 1201 CTATCAAAATGACATGACATTGCTTCTAATTAATTTGATTCAAGGTGAGACCTTACTTG 1260
QY 1261 attaaatgcttgcagaaagccttgaagaacagcaagagccttgatcacttgcctca 1320
Db 1261 ATTAATAGCTTGGCGAAGAGCTTGAAGACAGCAAGAGACTGATCACTTGCCCTCA 1320
QY 1321 gaaaagttattatgatatagaatccacacacacacacacacacacacacacacacacac 1362
Db 1321 GAAAAGTTATTTATGATAGTAATCTCAGTACACATCATATGA 1362
RESULT 2
AF132035 3435 bp mRNA PRI 08-APR-2000
LOCUS Homo sapiens core 2 beta-1,6-N-acetylglucosaminyltransferase 3
DEFINITION (C2GNT3) mRNA, complete cds.
ACCESSION AF132035
VERSION AF132035.1 GI:7527463
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 3435)
AUTHORS Kessel, A.G., Fukuda, M. and Clausen, H.
TITLE Control of O-glycan branch formation. Molecular cloning and
characterization of a novel thymus-associated core 2 beta1,
6-n-acetylglucosaminyltransferase
J. Biol. Chem. 275 (15), 11106-11113 (2000)
JOURNAL MEDLINE 20219156
REFERENCE 2 (bases 1 to 3435)
AUTHORS Schwientek, T. and Clausen, H.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-1999) School of Dentistry, University of
Copenhagen, Nørre Alle 20, Copenhagen 2200, Denmark
FEATURES
1.3435
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"

```



```

gene
1. 3435
/gene="C2Gnt3"
CDS
862..2223
/gene="C2Gnt3"
/codon_start=1
/product="core 2 beta-1,6-N-acetylglucosaminyltransferase
3"
/protein_id="AA063156.1"
/db_xref="GI:7527464"
/translation="MKIEKCYFKHTLQOKVFIELTLMLSLKLNVRRLFPQKDIY
LVEYSLSSTPEVRNRYTHVDEVRVENCISYOEPLIEIGKSLERRDIILEDDE
VYAMTISODIYQIOLRQYAKLVSKSEKSFPLAYELVYKDAIWEERLHAIVONHY
CIHYDRKAPDTEFKVANNKLAKCSNIFLASKLEVEVAHLSRLQDNCISDLKSSY
OMKXVINLGGDFPKNSNELSVSLKLNANMEFYKSPNSKREFFRYHSELRLRYVY
EYKPLIRTNISKEAPHNIOIFYGSAYFVLSQAFVYIINNNSIVDQPFMSKDTSP
DEHWATLIRFPGIIFGELSRSDVSDLOSTKTRVKNNTYEGFIFPCTSHLSVCI
YGAELFMWLIKDHWFANKFDSKVDPIILIKCLAEKLEEQRDWITLPSEKLFMDRNL
TTS"
BASE COUNT 1032 a 678 c 651 g 1074 t
ORIGIN

```

```

Query Match 100.0%; Score 1362; DB 9; Length 3435;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagataatcaatgttattttaacatacaccacagcaagaagtttcacccgttt 60
DB 862 ATGAGATATTCATTAATGTTATTTAAACATACCCACAGCAAGATTTCATCCTGTT 921
QY 61 ttaeccataagctctctcttctttaaagcttcaaatgtagagagaccccttcgcaa 120
DB 922 TTAACCCATAGCTCTCTCTCTTTTAAAGCTTCTAAAGTGAAGAGAGCTTTCCGCA 981
QY 121 aaagacattacttgctgtagactccctaagtaactgcgccttgtagaagaacagatrac 180
DB 982 AAAGACATTTACTTGTTGAGTACCTCCTAAGTACCTGCTTTGTAGAAACAGATAC 1041
QY 181 actcatgttaagatgaagtcagatgaagtaagtcgttcggtatcatatgaacaggag 240
DB 1042 ACTCATGTTAAAGATGAAGTACGATGATGAAGTAACTGTGGGATATCATGAACAGAG 1101
QY 241 ccttggaagatggaagaagtcgtgaagaagaagaagagacatcatgacttgagagat 300
DB 1102 CCTTGGAAATGGAAGAGTGTGGAATTAAGAAAGAGGAGCATGATGAGGAT 1161
QY 301 gatgatgttgygcaatgacagatgattgtagacattatccagactctaaagagttatgct 360
DB 1162 GATGATGTTGTGGCAATGACAGTATGATGATTAATACAGACTCTAAGAGGTTATGCT 1221
QY 361 caaagatgtgtcctaaagagagagaagagctcccaatagcctatcttctgtgtccac 420
DB 1222 CAAAAGCTGTCTCAAAAGAGAGAAAGCTCCCAATAGCCTATCTTTGTTGTTCCAC 1281
QY 421 aaagatgcaattatgctgaaagagcttccatgctatataccaacagacacataattac 480
DB 1282 AAAAGTGAATATGATGTTGAAGGCTTATTCATGCTATATACAAACACACACATTTTAC 1341
QY 481 tgcacatcatatgactgtaagagcactatataccttcaaatgtgocataaacttagct 540
DB 1342 TGCATCCATTAATGATCGTAAGGACACTGATACCTTCAAAAGTGGCCATTAACAATTTACT 1401
QY 541 aagtgcttcacatatttctgcttcccaatagagagctgtgagatagccacact 600
DB 1402 AAGTCTTCTCCATATTTTCATTCCTTCCAAATTAAGGCGTGTGAGATATGCCACATT 1461
QY 601 tccagactccagctgatttaattgctgtgcgaccttcgaagcttccaatccagtg 660
DB 1462 TCCAGACTCCAGCTGATTAATATGCTGTGCGACCTTCGAACTCTTCAATCCAGTGG 1521
QY 661 aaatgttatcaacttgctgtgcaagaatttccctcgaaatcaaatlttgaatttggt 720
DB 1522 AAATATGTTATCACTGTGTGGGCAAGATTTTCCCTGAACTCAAAATTTTCAATTTGATG 1581

```

```

QY 721 tcaagattgaaaaaactcaatlgagcaaatatgttggagacgg tgaaccccccaacagt 780
DB 1582 TCAGAGTTGAAAAAATCAATGAGAGCAAAATATGTTGAGAGCGGTGAACCCCCCAACAGT 1641
QY 781 aaatggaagatcacttaccatcatagaacttagaaggggtgcttatgataatgtgaag 840
DB 1642 AAATTGAAGATTCACCTTACCATCATGAACTTAGAGAGGGGTGCTTATGAAATATGTAAG 1701
QY 841 ctaccaaagagacaacatctccaaaggaagcaccoccccaataacatccagatattgtc 900
DB 1702 CTACCAATAGAGACAACATCTCCAAAGAGACACCCCCCATTAACATTCAGATATTTGTT 1761
QY 901 ggcagtgctatcttctttaaagtaagcattgttaataatatttcaaacactccatc 960
DB 1762 GCGAGTCTTATTTGTTTAAAGTCAAGCATTTGTTAAATATATTTTCAACAACACTCCATC 1821
QY 961 gtccaagacttlttgctgtgctcaagaacacatactctcgtatgagacactttgagct 1020
DB 1822 GTTCAAGACTTTTTCGCTGTGCTTAAGACACATACCTCTGATGAGCACTTTTGGGCT 1881
QY 1021 acctgattcggtgtccagaataactcgggagatltccagatccagccagatgtgtct 1080
DB 1882 ACCTTGATGGGTTCCAGGAATACCTGGGAGATTTCCAGATCACCCCGAGATGTGCT 1941
QY 1081 gatctcagagtaagactcgccttgctcaagtggatgaaatgaaagctttttctatccc 1140
DB 1942 GATCTGAGAGTAAAGACTCCCTTGTCAATGGAATTAAGCTTTTCTATCCC 2001
QY 1141 agttgactgatactcaacttcgaagcggtgtatttaagagctcgaagaattaagtg 1200
DB 2002 AGTTGACTGATCTCACTCTCGAAGCGTGTGATTTATGAGACTCGAGATTAAGGTG 2061
QY 1201 cttaacaagaatgagacatltggttgcataaattgattcgaagtggagccctactctg 1260
DB 2062 CTTTCAAAAGTGCACCTTGTGCTTAATTAATTTGATTTAAGTGAACCTTATCTTG 2121
QY 1261 attaatgcttggcagaagaagcttgaagaacagcagagagactgatacttgcctcca 1320
DB 2122 ATTAATAGCTTGGCAGAAAGCTTGAAGAACACAGAGAGACTGATCACTTGGCCTCA 2181
QY 1321 gaaagctattatgataagaactctactaacacatcatga 1362
DB 2182 GAAAGTATTTATGATGAAGAAATCTCAACACATCATGA 2223

RESULT 3
AC093259/c
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone RP11-229c3, WORKING DRAFT SEQUENCE,
6 unordered pieces.
ACCESSION
AC093259
VERSION
AC093259.1 GI:15193393
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 184590)
AUTHORS
DOE Joint Genome Institute.
TITLE
Sequencing of Human Chromosome 5
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 184590)
AUTHORS
DOE Joint Genome Institute.
TITLE
Direct Submission
JOURNAL
DOE Joint Genome Institute.
COMMENT
Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information

```

Center Project Name: 494352
Center clone name: RPCI-1L_229C3

Summary Statistics

Consensus quality: 176980 bases at least Q40
Consensus quality: 182494 bases at least Q30
Consensus quality: 182978 bases at least Q20
Estimated insert size: 183600; agarose-ef estimation
Estimated insert size: 184090; sum-of-ctnigs estimation
Quality coverage: 8.93 in Q20 bases; agarose-ef estimation
Quality coverage: 8.9 in Q20 bases; sum-of-ctnigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2136: contig of 2136 bp in length
* 2137 2236: gap of unknown length
* 2237 5820: contig of 3584 bp in length
* 5821 5920: gap of unknown length
* 5921 8425: contig of 2505 bp in length
* 8426 8525: gap of unknown length
* 8526 42924: contig of 3439 bp in length
* 42925 43024: gap of unknown length
* 43025 98717: contig of 55693 bp in length
* 98718 98817: gap of unknown length
* 98818 184590: contig of 85773 bp in length.

FEATURES

SOURCE

1. 184590
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-229C3"
/clone_id="RPCI human BAC library 11"

BASE COUNT 56307 a 36971 c 36814 g 53956 t 502 others
ORIGIN

Query Match 99.8%; Score 1358.8; DB 2; Length 184590;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagatatacaatgttatatttaaacatccctacagcagaagatttcacccgtt 60
Db 22472 ATGAAGATATTCAAATGTTATTTAAACATACCCCTACAGCAAGATTTCATCCTGTTT 22413
QY 61 ttaacctatggtctctctcttctttaaagcttctaagtgaagacatcttcgcgaa 120
Db 22412 TTAACCCATGAGCTGCTCTCTTTGTTAAAGCTTGAATGAGAACGACGCTTTCCGCA 22353
QY 121 aaagcattactggtgtgaagtaacctaaagtaacctgctcttgaagaacagatac 180
Db 22332 AAAGCATTACTGTTGGTGAATACCTCCTAAGTACCTGCTTTGTAAGAAAGATAC 22293
QY 181 actcatgttaagatgaagtcaggtatgaagtaagtaagtaagtaagtaagtaag 240
Db 22292 ACTCATGTTAAGATGAAGTACAGGTATGAAGTTAACTGTTGGGTATCTATGAACAGAG 22233
QY 241 cctttggaatgtgaagaagctctggaataaagaagaagagacatcattgacttgaagat 300
Db 22232 CCTTTGGAATTTGGAAGAAGTCTGGAATAAGAAAGGAGCATCTGACTTGAGAGAT 22173
QY 301 gatgatgtgtgacacatgacagatgattgacattacacagactcacaagatgattgct 360
Db 22172 GATGATGTTGTGGCATACACAGTGTGATGATTTATGAGATCTTAAGAGGTTATGCT 22113
QY 361 caaaagcttctcacaagagagaaagcttcccaatagactatcttggattgttcac 420
Db 22112 CAAAAGCTTGTCTCAAAAGAGAGAAAGCTTCCCAATAGCCTATCTTGTGTTGTCAC 22053

QY 421 aagaatgcaattatgttgaagaagcttatccatgctatatcacaccagacatatattac 480
Db 22052 AAAGATGCAATATATGTTGTAAGGCTTATCCATGCTATATACCAACGACCATATTTAC 21993
QY 481 tgcattcattatgatcgttaagacacgttatcctcaaaagttgcatagaacatttagct 540
Db 21992 TGCATCATATATGATGCTATGAGGACCGCTGATACCTTCAAAAGTTCCATGAAACATTTAGCT 21933
QY 541 aagttcttcccaataatttcatctgttcccaattgaagagcgttgaataatgccacatt 600
Db 21932 AAGTCTTCCCAATATTTTCATTTGCTTCCAAATGAGAGCGTGGAAATATGCCACATT 21873
QY 601 tccagactcagaagctgaatttaatttgcctgtgagacaccttcgaagcttcaatccagt 660
Db 21872 TCCAGACTCAGGCTGATTTAAATTCCTGTGGAGACCTTGTGAAGCTTCATCCAGTGG 21813
QY 661 aaatatgttcaactctgttggcagaagatttccctgaagtcacaatttgaattgtg 720
Db 21812 AAATATGTTATCAACTTGTGTGGCAAGATTTCCCTGAAAGCAATTTGAATTGGTG 21753
QY 721 tcaagatgtgaaaaaactcacaatgtagcaaatatgttgaagcaggtgaaccccaaacgt 780
Db 21752 TCAGAGTTGAAAAAATCAATGAGCAAAATATTTGGAAGGTTGAACCCCAACAGCT 21693
QY 781 aaattggaagattcactcactcactcactcactcactcactcactcactcactcact 840
Db 21692 AAATTGGAAGATTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 21633
QY 841 ctcccaataagagcaaacatctcagaagaagaccccccccaataacatgaatttgtt 900
Db 21632 CTACCAATTAAGCAAAACATCTCAAGAGAACCCCCCAATTAAGATTTGTT 21573
QY 901 ggcagtgctatttcttgaagcaagcattgttaataatatttcaaacactccatc 960
Db 21572 GGCAGTGCTATTTTGTGTTAAGTCAAGCATTTGTTAAATATTTTCAACAACCTCCATC 21513
QY 961 gtccaagacttttgcctggtcttaagaacacactcctcctgaagacactttgggct 1020
Db 21512 GTTCAAGACTTTTTCCTGCTGCTTAAGACACATCTCCGATGAGCACTTTGGGCT 21453
QY 1021 acctgattcgggttccagaagaatcctggggaatttccagatcagaacagatgtgct 1080
Db 21452 ACCTGATTGCGGTTCCAGAAATACCTGGGAGATTTCCAGATCCAGCCAGATGTGCT 21393
QY 1081 gactcagaagtagaactgcctgttcaagtgaattactactagaagcttttctatccc 1140
Db 21392 GATCTCAGAGTAGACTGCTGCTTGTCAAGTGAATTAATAGAGGCTTTTCTATCCC 21333
QY 1141 agttgactgagatcaccctcgaagcgtgtgatttcaagagctcgaagtaaggttg 1200
Db 21332 AGTTGTAAGTAGATCTACCTTCGAACCGTGTATTTATGAGAGCTCCAAATTAAGGTGG 21273
QY 1201 ctatcacaagaatgacattggtttgtctaaatttgatttcaagttgaacctatcttg 1260
Db 21272 CTTATCAAAAGATGACATTGTTGCTAATTAATTGATTCTAAGTGAACCTTATCTTG 21213
QY 1261 attaaatgcttgcagaagaagcttgaagaagcagaagagatgatacatttgcctca 1320
Db 21212 ATTAATGCTGCGCAAAAAGCTTGAAGAACGAGAGAGATGATCTTTGCCCTCA 21153
QY 1321 gaaaagtattatgagataagaactcactaccacatcga 1362
Db 21152 GAAAAGTTATTATGATAGATAAATCACTACCATCATATGA 21111

RESULT 4
AB037596 3508 bp mRNA ROD 20-Oct-2000
LOCUS
DEFINITION Mus musculus Ignt B mRNA for
AB037596
ACCESSION
VERSION AB037596.1 GI:9650953
KEYWORDS beta-1,6-N-acetylglucosaminyltransferase B.

Db	Accession	Length	Source
Db	1531 CTG6TGGCCTCCAGATCCCTGAAATATGCTCTCAACACCTGCGGAGCACTTTTCT	1590	GenBank
Qy	697 ctgaagtcgaatttgaattggttgctcagagttgaaaaactcaatggaacaataatgttg	756	GenBank
Db	1591 CTGAAACCAACAAAGAAATAGTTGACAGTCTGAAACCGTTATATGGGAAACCTCAC	1650	GenBank
Qy	757 gagaacggaaccccccaacagtaaatgtgaagaatcattcattccatcagaacttga	816	GenBank
Db	1651 CCAAGGCGTCTGCTCTCGCACATGCACTTGGAGAACCAAGTACGTCACCAGAGCTG	1710	GenBank
Qy	817 cggatgaccttgaatatgtgaagctacaataaaggaacaacatctccaaggaagacc	876	GenBank
Db	1711 TTAGNCCATATAAATCCCTACGTG-----ACATATCAGCAGATTTAAAGCTCCCC	1764	GenBank
Qy	877 ccccaatacatcagatatgttgctggcagtgcttaattgttttaagtcagaactgtt	936	GenBank
Db	1765 CCTCACAACCTGACCATTTACTTTGGCAGCTGCTATATGCGCTCAGCACAGGAAATTC	1824	GenBank
Qy	937 aaataatlttcaacaactccatcgttcaagaacttttgcctggtctaaagaacat	996	GenBank
Db	1825 AACTTTGGCTCAAAAGACACAGCTTCACTAGACTTAATCTCTGTGTCGAACAGACGT	1884	GenBank
Qy	997 tctctgtatggacattttgggctaaccttgatttggtgtccaagaatacctgggagatt	1056	GenBank
Db	1885 AGTCTGATGAGCACTTGTGGGTGACACTCAATAGACATCCCTGGAGTCTCGCTCC	1944	GenBank
Qy	1057 tccagatcagcccaagatgtgtctgattcgtgaagtaagtaagctgcctgttcaagtga	1116	GenBank
Db	1945 CCACCAACGCGATCTGGACGCGTAACCTCAGAGCTGGAAGTGATGAGATGGAAACG	2004	GenBank
Qy	1117 tactatgaagcctttctatccagtgtaagtgatctcacccttggaaggtgtgatt	1176	GenBank
Db	2005 AAGCATGAGCG-----TGCATGTGCTACTACGTCACATGGCATTTGTATC	2049	GenBank
Qy	1177 taatgaagctcagaattaaggtggtctatcaaaagtgaatgtggtgtcataaaatt	1236	GenBank
Db	2050 TATGGAACGGAGAGCTTCAGAGTGGTGAATTATTTCGCAAAACCTGTTGCTACAAATTT	2109	GenBank
Qy	1237 gatctaaagtgggccctatcttgattaaatgagcttgagagaagaagctgaagaa	1290	GenBank
Db	2110 GAACCAACACATACCCCTTACCGTGAATGCCCTGGAGACAGAGGCTTCGAGAA	2163	GenBank
RESULT	5		
LOCUS	AF318573	108873 bp	DNA
DEFINITION	Bovine herpesvirus 4 long unique region, complete sequence.		
ACCESSION	AF318573	AF271211	
VERSION	AF318573.1	GI:12802528	
SOURCE			
ORGANISM	Bovine herpesvirus 4.		
	Bovine herpesvirus 4.		
	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae.		
REFERENCE	1 (bases 1 to 108873)		
AUTHORS	Zimmermann, W., Broll, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and Goltz, M.		
TITLE	Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and Identification of an Origin of DNA Replication		
JOURNAL	J. Virol. 75 (3), 1186-1194 (2001)		
PubMed	11152491		
REFERENCE	2 (bases 1 to 108873)		
AUTHORS	Zimmermann, W., Broll, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and Goltz, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-NOV-2000) Project P24, Robert Koch-Institut, Nordufer 20, Berlin D-10555, Germany		
FEATURES			
Source	1..108873		
	/organism="Bovine herpesvirus 4"		
	/db_xref="taxon:10385"		
	misc_feature 1..108873		

CDS

/note="long unique region; LUR"
 complement(445..633)
 /note="ORF B01"
 /codon_start=1
 /evidence=not_experimental
 /product="hypothetical protein pb01"
 /protein_id="AAK07921.1"
 /db_xref="GI:12802539"
 /translation="MEGDVSNVAFGRKRTQDNLNNSAGNEVCCAVLREKNNVNSAGC
 CAVKRNNSFHSILPGMP"

CDS

614..4480
 /note="tegment protein; ORF3 B0F0A1"
 /codon_start=1
 /product="r-fgAM-synthase"
 /protein_id="AAK07922.1"
 /db_xref="GI:12802530"
 /translation="MDTSPSTQLCFVYSRAPLNSNECVIQQTSDGQLFISVPLQY
 EGVFVAVNGEQAVERLVPDLKLFSSHGLPFFHSSILKIDSRLSHYGVSPSA
 PETAISRLVAILQSMGLAPMSNLTOILKIVLSIISIKCNLPSPKIMHSLRAL
 CPDLSLFRPKYSEIVGAHETGAPGVLSPPKGLLGAFAVGGLNAGSAGGFMT
 POGPOTIVNQNFCKTAYIVDVWYIPEISYTLSSGHMTDGLAYVGOLOYAPRD
 LKLVNSPNMGVNSPVSSQIGASTVSTTSPRSSVELPYTRDMDVAYMNRCC
 SLPTQCTEPPIVESHPLDVSPECPAFIAGAVCFSRNITTDLPPEIRKINIQGLT
 FTSCHDSGREVRCILGALKIMSFLEPSQYKISAVKCEIMHAKILVCOQYIGIK
 IYQALPPPLNTGLLPFTASNROQNMIMIKQHLINIVAPVLVAIEATPNTTDYVLEK
 AGCAHGCQYVIGRTRQPHIHYKDKQDNVERKISLKRSPSYSLSTNKLQKOS
 ENMODSSEMDLPSLSTYLLIHKFAVCKDFIVKHIDRLSSGRAVOOQIGARDI
 PISDSIIVSDLSLAAPDGSENPMATASLDDLPVNNPEVPGICSAIGCEPFLS
 TYPPIKGAVHVAIVESLNIIVASPKRIDDTVCPTNITPOTDSHGKILNAGRE
 CSOLGICVFTSCSTSNRQKCLNLSIVRLYLTATAPKDYQGLTPLKPPDSIT
 VMLPISTHYHFRGIVISQIFRDSPSGHTINDPIYVKKLITVTKHLNHSITSCHY
 GCGGLITACFEMAYAGGASIALIVPOBDDPYLFTLSTPGLMEVPRKVSIVQKHE
 TSDIIVDVGRPLPSVANSNTFVSYKRRIIFRESLNEMVENWHRFSTKEDMRYPCY
 QGTEDPDKLQHLITFQPSMCPHGPYKHOVANYLLPGTNPDSLVALEEGFRN
 LVSTFTDKTVLIVDTNVEIGLIGANILEDALIGKAIMYKNSVYALAEAGFKRI
 NSPVEISLAIQACQILFENKFMGYNKSPNTMYCKENSGMESRLNAGELFKLI
 AVALOSNGKSLPGMIGSTHGFAPHSBETVEMSTHGMAVATOCYAGDISAPLTP
 ONPTAGYVLSIGCSADGRHLALHDPLCSNLMQMPHI PKMTPLKYSPKRMFLDLH
 IWANKVRMDPPPPHPDLNRYM"

CDS

4608..4811
 /note="ORF B02; B0F0A2"
 /codon_start=1
 /evidence=not_experimental
 /product="hypothetical protein pb02"
 /protein_id="AAK07923.1"
 /db_xref="GI:12802531"
 /translation="WASPTQYCHMVRVYHGRKSAKISYGVKDRDYHQLTSKNSAYS
 GMSPPCPCADLNLKLFVSQCPIS"

CDS

5286..5513
 /note="ORF B03"
 /codon_start=1
 /evidence=not_experimental
 /product="hypothetical protein pb03"
 /protein_id="AAK07924.1"
 /db_xref="GI:12802532"
 /translation="WGFNPVMLYFHPCCNCTPTSLKHGLYIKRYISYLIRATFLVRNV
 HIGPKRDPTRPPGPGPSILSLVQRLAKK"

CDS

5943..9347
 /note="ORF 6"
 /codon_start=1
 /product="single-stranded DNA-binding protein MDPB"
 /protein_id="AAK07925.1"
 /db_xref="GI:12802533"
 /translation="MAKTAQOQOQYEDNIGSKAPICPGCYIYYKKREPTKEASLLG
 NGEPTSVSWYDLGLTVEGEPFLNKAHAKIDITTVSVKVSYHREPMFENNIC
 EPIFHGGLPOLQOETRDILGFTHVSPSPNTNNISLQDLCGDNCDDEWILLVAT
 EGFRERLYHGHKIPILGHVEVOISPGHAARKPIYEDDLCSKTAIPRMKFRHVDYS
 QYLENSLTAHQALRYKDVATVIOAMSSQVVRDGHKPKVYHKKDPSSSSRPGDGL
 SLIMDSVASELAVSYGLSFLIDAPQDMSALLDYTSWPLFVDCITIEERLQALHAWNK
 OSIHNATOLEFNSMLYLRIOKQNPKAQGLDLYNNSVYLQHGISTYSEATQOYEGQ
 PVFOGQGNLSGTYITLHLLAASMCPTNLARCYLFCOCHQSTONGSYNISY
 VQTAASDMCNLCQGRKPAVCINLTFEYRLRDRPEIITSHRPDIYIINGSGMNDID
 ILGNFASFREKDEGAQVEYQKTYTWMQLTQTLLEKLEGMGINKTTPPEKLYVSDIPSP

gene

11420..14044
 /gene="g8"
 /note="g8"
 /codon_start=1
 /product="glycoprotein B"
 /protein_id="AAK07927.1"
 /db_xref="GI:12802535"
 /translation="MYKTLIFPALIKVCSFNQTTHTSTTSPSISSTSTTSTSK
 PSNTSTNSLASPONTSTSKSPSTDNQSTSTPTIPTYDIDTASNRKTRKCASSS
 SGLTFRFDLQTCPTDRKRAHEGILLVAKNIVYIRKRYRKAIVSVYRWSSQ
 AAVTNRDIDISRAIPNEISIMDIRHAYCFAMKAVNGIANTYIDDSBNKSVLPQVA
 GLENTENRFSOPLIYAEDEGPPGIYRVATTNCEVDEYVARSYPYHIFALQDPI
 EISPECHNNSOCTGNSSTRDAKTVIENHODTVYERGHGTRKRLIKLEDEYIS
 WKADEBRALCDPVTKTPPRAIOTIHNSPFRVANEVASLTSNOEETLRGTEI
 LKCMSTINETLEETVYKRNKSHIDGEEKYKTKNGGLFLINQAKKPLNLSHTVYTI
 ERNNKSTNSRQKRSYDTKFOGAKGLSTAOVQVYVDHRTISMNHLLEHLITKWRQ
 KQDNLWYELSKINPVSAVAIYAGKAVAKAGDAFVWSECINVOASVNIHRSKMTY
 DKPVQYSRPLVYERFVNSTATFEGQLGTENEILLNTHVETRPADAHFVYKNMTHY
 EKYKFKVMDTNNISTLDTEFLNLTFDNIDFKVELYSETERKMASALDETFGR
 EYVYVQKULASLBEDLNTIDLRDLVLDLSMAADLGDIGKVVNTFSGIVTEGFS
 INGVFVSFEFNPFGVYTIILLIIVYFVYFVYSRNNNENAEIIMYIENIDKASEOE
 NIQPLGEEIKRILLGMHQLQOSEHKSSEBASHRGLFOLLGDLQILLRRGYTRLP
 TFDSPSNDTSETHOKYV"

gene

14184..17201
 /gene="DPO1"
 /note="DPO1"
 /codon_start=1
 /product="DNA polymerase"
 /protein_id="AAK07928.1"
 /db_xref="GI:12802536"
 /translation="MEFNPYIYSKKKTKDRKISGVDDPDKTNPYCALYKCRKPTGPT
 GVSITTPDPLVLEKDEKPIPTSGDHSRYSAKAPAPAPPEKDKLFFVYDIE
 TITPDRCPDPIPOFOTDILPNTVTKILGKQDQASVGVNYYTQASAPVEINQVGN
 LSVQOCTINGVYNNKCTCKFSITQAEKLIKIDEPDLSVYVYITLSSAPVEINQVGN
 ISGCEQTESNVANSTRFTIINKFTSTFGVYSCGNPPRISQDRSDRLDEPDCGLDLO
 FHEDEQQMPYTTMSDIEDICIGQGPCKATKDEDTLQISLIITWYCSPPONATILS
 LGQCDIANDYVPESELDMEYAFPTLLARDNFIDVTYNNIANDEFPPIIDRQAVY
 NITIRHFTKISGTYSEVNOQPMNTGQAFKRSKISIGSLVDSYDVMQVLYDKMLISDY
 KLVNTFAKHCQGTQKEDYVYKEISLFRSEAGARIGSVCYSDVYVLDLLKFTMAYH
 EISEIAKMAKIPARVLTQGGQIRVESCLEBAKAKRHFLIPGAGKPGQYQATYIN

CDS

9352..11430
 /note="ORF 7"
 /codon_start=1
 /product="transport protein"
 /protein_id="AAK07926.1"
 /db_xref="GI:12802534"
 /translation="MAREIAVYSQVFDLSERTSLITFCDDPHDLGAFQNNKRLKL
 LIDALVSHLVEONKRESCSLLEHLINITYITMQLLETOLNENYERKDFAMLHLS
 DICCFKRYNLFGDYCVPLISLINDIIEIFSNLSNVFVICSNNALDELETINF
 LGLRLRISPIPLDYLPEISCMQNLNETSILPNGENVFTIMNSINCTHATVAPR
 POGLENEMLHLNIPVQDTRKVOLESDRFLQASIEALDOHIFEKASHOLEIS
 NLIYVNSGHSNDOSPKLSQSQMATLLOHETRMAHROKVASHLATPPRIHPSVRA
 YPESLFCGGITYSINDNQAOKQCSSTFLAKSNKYMIRKNTLYVRNLNLAHSP
 IREGDIDRPNISIKQARHTIGETSEQIFSDAKRKNNTAQKTTEKITTQCLET
 QSKLVYQKLSQSHIDSITLQFVSLITGPLEGDPPEPSPNILLANCLDAKVMFHHK
 MLISEYFMSMEPKDMIDAEFAFNINGKTLNVSQCAWRYIRELVSVSLVYNWME
 KQDNIYESTLSPHNLFSPPKFGKGIYVETKTAFLPIFIHNTGMIRKDYALYHH
 LQTSKKNVL"

CDS

PIRFGYNTPLVYFASLYPSIIQAHNLCTSTMIDQNLHLHLPDDETEHLSIGP
 IHFKOHKHTSLSTILLTAMARKTIRKELANCDGPKRTILDKOOLIKYTCNSVY
 GETVASGILCPIDIAETVTLQGRMTLEKSELANCDGPKRTILDKOOLIKYTCNSVY
 RYVIGDTSLDIECRGYPMSOVLNFPADKLETTETTKLEKXEPFOLLMTK
 KRYIGLSSKXIKMKGVDLIRKACSFVOTTEITLIDLPREPKYQAOYICRCRAVA
 KYSDSLPVGFKVYIDVLYNNSYALATGKVSBSLSTELSLPREYKFTTLPLHTV
 YRKINSRNEELPQIDHRIPIYFKDQKCKSDLPYVSQNKIPISLVEYFKLI
 HGVAINQCLLEFGNSNMVTEILYNFNIVISLT"
 17297. 18577
 /note="ORF 10: B0FB1; conserved in other
 gamma-herpesviruses"
 /codon_start=1
 /product="unknown"
 /protein_id="AK07929.1"
 /db_xref="GI:12802537"

CDS

complement(18848..19345)
 EDNMVATLGENIYSRCHLNKSENTPSPMDTP"
 /note="ORF B04; short ORF of immediate early transcript 1
 RNA"
 /codon_start=1

Query Match 14.3%; Score 194.8; DB 14; Length 108873;
 Best Local Similarity 52.7%; Pred. No. 7.9e-37;
 Matches 520; Conservative 0; Mismatches 427; Indels 39; Gaps 3;

QY 316 atgagcagatggttgacattatcagactctagaggtatgctcaaaagctgtctca 375
 Db 107637 ATAAACGAGACTGTGAGCGCTTCAAGGCCAAAGAGATTCAATACGTTCCACATGAGT 107696
 QY 376 aagagaggaagaagctcccaatagcctatcttggctgcacaaagaatgcaatag 435
 Db 107697 AAAGAAAGAGCTAGCTCCCATGCTGCTCATGATGCTGATGAGAGATTGAAGAAC 107756
 QY 436 gtgaaaggtcttcacgtctatatatacaacagacaatttaccgacccattatg 495
 Db 107757 TTGTAAGCGCTGCTGCAAGCGGTGTATGCCCCGAGACATATATCTGTGCTCCAGTGGAT 107816
 QY 496 cgttaagacacctatagctcaaaagtgtgcataaacaatttagctaaagtctctccaat 555
 Db 107817 GTGAAAGTCCCGAGACTTCAAGAGAGCGGTCAAGCGCATATATTGCTGCTCCCAAT 107876
 QY 556 atttcattgctccaataatagagctgtgtgaatatgcccacattccagaccagagct 615
 Db 107877 GTCTTCATGSCAGTAAAGTGTGTCGGTGTATATGCTCCTCGTCCAGAGTGAAGCT 107936
 QY 616 gatttaaatgtctgtcgaacctctgaagcttcaatccagcaggaataatgttacaac 675
 Db 107937 GACCTGAACGTGTATGGAAGAGCTGCTCCAGAGTCAAGTGCATGGAAGTACTTACTGAAT 107996
 QY 676 ttgtgtgtggaagagatttcccccgaggtcaaatttggaattgtgtgcagaagtgtgaataa 735
 Db 107997 ACAATGCGGACAGACTTCCCATTAAGAACCAATGCCAGATGCTCTGCGCCCAAGATG 108056
 QY 736 ctcaatgagacaatatgttggagagcgttgaaccccccaacagtaaatgtgaagattc 795
 Db 108057 TTGAAGGTTAAGAACATATGAGAGTCTGAGCTTCTGATAGTCCAAAAAATTCGCTG 108116
 QY 796 acttaccataatgaacttaagcgggtgccttaagataatgtgaagcttaccataaggaac 855
 Db 108117 AAAATACCGCTATAGTGAGAACACACTGTACCTT-----ACC 108155
 QY 856 aacatcccaagaagaagaccccccaatacatcatatattgttgcagtgcttattt 915
 Db 108156 AGCAAGATGGAAGACCTCCCTCCGATTAATTACCCATGTTCCACAGGGAATGCTCTATT 108215
 QY 916 gtttaagtcagcaattgttaataatatttcaacaactccatcglttcaagacttttt 975

Db 108216 GTGAGCCTCTGAGACCTTTTCCAAACATGCTTAGAACAAACCTTAANCCAAAGACTGGTT 108275
 QY 976 gctgtctaaagaacacacacactctctgtatgacacttttggctcaacttggatttggatt 1035
 Db 108276 GAATGGTGCAAGACCCCTATAGCCCCGAGAACACCTCTGGGCCACCTTCAGCGTGCT 108335
 QY 1036 ccaggaatacctcgggagagattccagatcacccagagatg---tgtcatctgcagagt 1092
 Db 108336 CCGTGATGCTCGTGGCTGCTGCTTCTGATGCGCACCCCAAGTATACATCTCAGACATGACTGCC 108395
 QY 1093 aagactcgcctgtcaaatgtgaatttactatgaagcttttctc-----tat 1137
 Db 108396 ATCGCCAGGCTCGTCAAGTGGGAGTACCAACGAGGAGATGTCAGCATGGGGCGGCTTAT 108455
 QY 1138 ccaggtgttaccgtgactcaccctcgaagcgtgtattatgagctgcagaataag 1197
 Db 108456 GCACCGTGTCTGAGATCATGAGAGGCGCATCTGATTCAGGGCGGGAGCTGTAC 108515
 QY 1198 tggcttatacaagaatggaacttggcttgaataaattgatttcaagtgtgaacctatc 1257
 Db 108516 TGATCTCTCAAGACCATCACCCTCTGGCAACAAAGTTGACCCGAGGCTGATGATATAC 108575
 QY 1258 ttgattaaatgcttgcagaagaagct 1283
 Db 108576 GTCTTCAGACTGCTTAGAAGATGACT 108601

RESULT 6

AB037597 1206 bp mRNA ROD 20-OCT-2000

LOCUS beta-1.6-N-acetylglucosaminyltransferase A, complete cds.

DEFINITION beta-1.6-N-acetylglucosaminyltransferase A, complete cds.

ACCESSION AB037597.1 GI:9650955

VERSION beta-1.6-N-acetylglucosaminyltransferase A.

KEYWORDS Mus musculus cell_line:F9 cell cDNA to mRNA.

SOURCE Mus musculus

ORGANISM Mus musculus

REFERENCE 1 (sites) Chen, G., Y., Kurosawa, N. and Muramatsu, T.

A novel variant form of murine beta-1, 6-N-acetylglucosaminyltransferase forming branches in

glycobiology 10 (10), 1001-1011 (2000)

2 (bases 1 to 1206) Kurosawa, N., Chin, G. and Muramatsu, T.

Direct Submission Submitted (25-JAN-2000) to the DDBJ/EMBL/Genbank databases.

Nobuyuki Kurosawa, Nagoya University School of Medicine, Department

of Biochemistry, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550,

Japan (E-mail: kurotsuru.med.nagoya-u.ac.jp, Tel: 81-52-744-2063,

Fax: 81-52-744-2065)

FEATURES Location/Qualifiers

1..1206 /organism="Mus musculus"

/db_xref="taxon:10090"

1..1206 /cell_line="F9 cell"

1..1206 /gene="Tgnr A"

/gene="Tgnr A"

/product="beta-1, 6-N-acetylglucosaminyltransferase A"

/codon_start=1

/protein_id="BAB03496.1"

/db_xref="GI:9650956"

/translation="MPSGVRFYFVSYTVYIVLVVLSRGGDSYOKINDSYVL

SOVSSPFDGSRFIMRKMLHKEKPSCTEYVDSHTTLPSEDFPFLAIVYH

HNFDIFALFRALFMPONTICVHDEKATAFKFAVQLVSCFPNATLASKMEVYIG

GISRQADLNCIKDLSTSEVPKVAINTCGDFLKNKEIVQILGKLKKNLTPGLV

PPAHAIGRVYVHREHLSKELSYIVRTALAKPPPHNLITVFGSAVVALSRFANFVL

Fri May 3 10:57:51 2002

us-09-645-192-1.rge

Page 8

BASE COUNT	300 a	318 c	285 g	303 t
ORIGIN	RDPRRAVDLLHMSKDTFSPDDHFWTLIRIGVPSGSPMNPMSWTGNLRVAKMMDEAHH GGCGHGYVHG-CIYAGNDQLQMLNSQSLFANKFELPNTYPLTVELELRRLPERTLNQSE IATQPSMYF"			

Query Match	14.3%;	Score 194.4;	DB 10;	Length 1206;
Best Local Similarity	52.1%;	Pred. No. 8.3e+37;		
Matches 497;	Conservative	0;	Mismatches 436;	Indels 21;
			Gaps	2;

QY	337	tataagcctcaaaaggttatgtctcaaaagctgtgtctcaaaagaggaagaagcttcca	396
Db	223	TATGTGACCCAAACCCACTATATCACTGACGGCCCCCTTTATCTGACGAAGAGGTGATTTTCT	282
QY	397	atagccattcttgtgtgtcccaaaagatgcaattatgtttgaaagctttccatgtc	456
Db	283	TTGGCGATGTCAATGGTCATCCATCCAAATTTGACACCTTTGCAAGGCTCTTCCAGGGCA	342
QY	457	atacaaacagcaaatattctgcacatccattatgacscfagaagcaactataccctc	516
Db	343	ATCTCATGCTCTCAAAATATCTACTGTGTACGTGGATGAAAAGCAACAGCTCAATTTC	402
QY	517	aaagtgcctatgaaacaattagctaaagtcttcctccaattttcattgtctccaatt	576
Db	403	AAAGTCCCGGGAAACGTTTAGTAGCTGTGTTCCCAATGCTTTCTGGCTTTCTAAGATG	462
QY	577	gagagcttggaaatagccacatttcgaagctcgaagctcagctgattaaattgtgtgcag	636
Db	463	GAGCGGGGGGTATGTGTGAATCTCCGGGCTCCAGGTCTGAOCTGACATGCATCAAAAG	522
QY	637	cttcgtgaagcttcaatccacagttggaatatgtataacttgytgygcaagattttcc	696
Db	523	CTGTCCACCTCCGAGGTCCTCGAATCGGATGAGCCATACACCTGTGGACAMACCTTCCC	582
QY	697	ctgaagcaaatattgtaattggtgtcagattgaaanaactcaatgagcaatatgtg	756
Db	583	CTTAAACCAACAAGAGATAGTTCACTACCTGATAAAGGCTTAAAGGGAAGAACCCTACT	642
QY	757	gagacggtgaaaccccccaaacagtaaatattgaaagaattcaattaccatcgaatga	816
Db	643	CCGGGGGGGTCTGCTCCAGGCCACGACCATTTGGAAGACACAGTATC-----TTCACACGG	696
QY	817	cgggttcctatgataatgtgaaagtcacaaataagaacaacatcccaagaagcacc	876
Db	697	GAACACCTTAACCAAAAGAGCTTCTCCTACGTATAGAACCAACAGCGCTGTGAAGCCCCACT	756
QY	877	ccccataacatcagatatattgttggcagtgcttatttgytttaagtcgaagcattgt	936
Db	757	CCCCAACACCTCACATTATTTTGGCTCTGCTATGTGGCCCTGTGCGAAGAGTTTGT	816
QY	937	aaatatatttcaacaactcatcagatggttcaagacttttgcgcgtgtcaagaacatc	996
Db	817	AACTTTGTTCCGTGACCCACAGGGCGGTGATTTGCTCATGTGTCAAAATATCTTC	876
QY	997	tctcttgatgagcaatttgggtctactctgatttcggtgtccagaatctccttggagat	1056
Db	877	AGTCCCGATGACATTTCTGGGTGACCTCAATATGAGATTCCAGAGTCCCTGACTCCAG	936
QY	1057	tccagatcagcccaagatgtgtcgtacgtcagagtaagaactcgtgcttcaagtgaat	1116
Db	937	CCACCAAAACGATCTCGTAGGGGAACCTGACAGCTGTGAAGTGAACATGGAAGCG	996
QY	1117	tactatgaagcttttctatccacagttgacgtcagatctcactctcgaagctgtgata	1176
Db	997	AAGCATGAGGC-----TTGCCATGGTGACCTACCTCCATGGCAATTTGTATC	1041
QY	1177	tatgagctgtcagaataaagttgcttatcaaaagatggaacatggttgcataaat	1236
Db	1042	TATGGAACCGAGACTTGTGAGTGTCTATTAATTCGAAAAGCCTGTTTCTTAACAAATT	1101
QY	1237	gattcctaaggtgcagccctattctgattaatattgttggcagaagaagcttgaaga	1290

Db	1102	GACCTCACACATACCCTTTACCGTGGAAAGCCTGGAACTGAGGCTTCAGAA	1155
RESULT	7		
LOCUS	AX087948		
DEFINITION	Sequence from patent WO0114535.	PAT	17-MAR-2001
ACCESSION	AX087948		
VERSION	AX087948.1	GI:13396926	
KEYWORDS			

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
Homo sapiens	Elkayacota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 1317)	Schwentner, T. and Clausen, H.	UDP-D-acetylglucosamine: galactose-4(gb)1,3- n- acetylglucosamine-6(ga)-r/(gc)nc to galnnc -g(b)1,6- n- acetylglucosaminyltransferase 8, cgnt8
	Patent: WO 01/4535-A 14-01-MAR-2003;			Schwentner, Tilo (DK) ; Clausen, Henrik (DK)	Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .1317
CDS

```

```

/note="The gene of C26n12
/codon_start=1
/protein_id="CAC34676.1"
/db_xref="GI:13396927"

/translation="MTQMKRIQCOLHYIMALGCMYLATVALIKLSPFLKQDSHFT
RESQJICRMIILYNFLKLPARKRSINCSGPTVKGQELAVLAILNLEIYKKRREFT
YSLTRDCEHFKAERKFTIQPLKEKEVEPIASVNIHEINFEESLRVAVAPQ
CVHDEKSPETFEAVKAIISCPNFIASIKLVRAVNSRQADQINCDEDLQ
PMYFLTCGTDPPIKSNAMQVALKILNLRNMESEVPPKHEITMKVKEIYVLR
HLNKKKDDPPVNLIMETGNAYIARSLDFQJHLKNDKSDILEWKDYSPEHHI
TLQRARMPSPVNNHPRKISIDMTSLASRLKMGQGHGDDIDKAPVPCSSIHRAI
YGAQDLNMLQNLNHLKANKDFPVDDALQCLDELYLRYKAIYTEL"

```

BASE COUNT	359 a	307 c	322 g	329 t
ORIGIN				

Query Match	14.1%;	Score 191.8;	DB 6;	Length 1317;
Best Local Similarity	52.5%;	Pred. No. 3.6e-36;		
Matches 517; Conservative	0;	Mismatches 432;	Indels 36;	Gaps

Qy	317	tgacacagtgatttgcacattatccagatcctaagaagttatgctccaagaagcttgctccaa	376
Db	317	TCACACAGAGACTGTGAGCAGCTTCACAGGCTGAAGAGAAATTGATACAGTCCCACTGAGCA	376
Qy	377	agagaggaataaagctctcccaatacactatttttggtgtgcacaaagaatgcaatttbg	436
Db	377	AAGAAGAGGTGGAGTTCCTTATGCATCTTATGGTGAATTCAGAGAAATTGAAAAC	436
Qy	437	ttgaagaagcttatccatgctatatacaaacagcaccacaaatttactgcatccattatgac	496
Db	437	TTGAAAGGCTAGTGGAGACTGTGATGGCCCTCAGAACATATAGTGTCCATGTGGATG	496
Qy	497	gttaaggcaactgcataccttccaagaatttgcacagaacaatttgcctgaagctcttcccaata	556
Db	497	AGAAAGTCCCCAGAAACCTTCAAAAGGGGCAACGACATATTTCTTGCTTCCCAATG	556
Qy	557	tttccatgcttccaactatagaagctgtgagataatgcccacatttccagactccaagctg	616
Db	557	TCTTATATGCCAGTAAGCTGGTGGGTGGTTATAGCCTCGTGTCCAGGGTGGAAAGCTG	616
Qy	617	atttaaatgtgttgcgaacttcttgaagcttccaatccagttggaatataatgtatccaaact	676
Db	617	ACCTCAATGCGATGAAAGACTGTGCTCCAGAGCTAGTCCCTGGAAATACCTTCCGTGATA	676
Qy	677	tgttggggccaagattttccctctgaagtcacaatttgaattgtgttcagagttggaanaaac	736
Db	677	CATGGGAGCGAGACTTTCCTTAAGAGCAATGCCAGAAATGTGTCCAGGCTTCAGAGATGT	736

QY 737 tcaatgagcaaatatgttggagacggtgaaaccccccaacagtaaatgtgaaagattca 796
 Db 737 TGAATGGAGAAATAGCATGAGTCAAGAGTACCTCTTAACCAAGAAACCCGCTGGA 796
 QY 797 cttaaccatcagaacttgaacggtggtccttaataatgttaaacataaagacaa 856
 Db 797 AATATCACTTTGAGTGTAGAGACACATTAC-----CTAACCA 838
 QY 857 acatctccaaggaagaccocccataacatcagataattgttgcagtgctattttg 916
 Db 859 ACAAGAAGAGGATCCCTCCCTTATATTAATACATGTTTACAGGGATCCGACATG 898
 QY 917 tttaagtcagaacttgtttaaataatttcaacaacatccatcgttcaagaacttttg 976
 Db 899 TGGCTTCCCGAGATTTCCTCCAAACATGTTTGAAGAACCTTAATCCCACTGATG 958
 QY 977 cctgttcaaaagacacatactctcctgattgaagacttttgggttaacctgttggttc 1036
 Db 959 AATGGTAAAGACACTTAAGCCAGATGAACACTCTGGGGCCACCTCTACGCTGAC 1018
 QY 1037 caggaataacctggtggaatttccagatacagcccaagatg---tgtctgattcgaagta 1093
 Db 1019 GGTGATGCTGGCTCTGCTCTGTTCCCAACCAACCCCAATAGCATCTCAACATGACTTCTA 1078
 QY 1094 agactcgcctgttcaagtggaattactatgaagcttttttctatccaggt----- 1143
 Db 1079 TTGCCAGGCTGTCAAGTGGCGAGGCTCATGAAGAGACATCGATAGGCTCTCTTATG 1138
 QY 1144 -----tgtactggatctcaaccttcgaagcggtgtgtatttattgaagctgcagaatgaagt 1198
 Db 1139 CTCCCTGCTCTGATCCACACCGAGGCTATCTCGTTTATGGGCTGGGACCTGATAT 1198
 QY 1199 ggcctatacaaatgagatggttggtaataaatttgatcttaagttgaagttaccatct 1258
 Db 1199 GATGCTCTCAACCATCCTGCTGTGGCCACACACTTTGACCCCAAGGTAGATGATATG 1258
 QY 1259 tgattaatgcttggcagaanaagct 1283
 Db 1259 CTCCTCAGCTTGAAGATACCT 1283
 RESULT 8
 AF038650 1317 bp mRNA PRT 25-MAR-1999
 LOCUS Homo sapiens core 2/core 4 beta-1,6-N-acetylglucosaminyltransferase
 DEFINITION (C2/4Gnt) mRNA, complete cds.
 ACCESSION AF038650
 VERSION AF038650.1 GI:4511860
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 1317)
 AUTHORS Schiøntek, T., Nomo, M., Levery, S.B., Merx, G., van Kessel, A.G.,
 Bennett, E.P., Hollingsworth, M.A. and Clausen, H.
 TITLE Control of O-glycan branch formation. Molecular cloning of human
 CDNA encoding a novel beta1,6-N-acetylglucosaminyltransferase
 forming core 2 and core 4
 JOURNAL J. Biol. Chem. 274 (8), 4504-4512 (1999)
 MEDLINE 99143102
 REFERENCE 2 (bases 1 to 1317)
 AUTHORS Schiøntek, T. and Clausen, H.
 TITLE Direct Submission
 JOURNAL Submitted (28-NOV-1997) Dental School, Faculty of Health Sciences,
 Copenhagen University, Noerre Alle 20, Copenhagen 2200, Denmark
 FEATURES
 source
 1. .1317
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1. .1317
 /gene="C2/4Gnt"

CDS
 1. .1317
 /gene="C2/4Gnt"
 /note="core 2/4-Gnt"
 /codon_start=1
 /product="core 2/core 4
 beta-1,6-N-acetylglucosaminyltransferase"
 /protein_id="ABD21525.1"
 /db_xref="GI:4511861"
 /translation="MVQMKRLQQLHYLWALGCGYMLLATVALKLSFRLKQSDPHLGLES
 RESQSYCNILNYFLKLPKRSINCSGYTRGDQAVYQALINLNEVKKREPFDTH
 YSLRDCHEFKAEKRFLOEPLSEKEVEEPILYSVNIHEKINERFLRAYAPONIY
 CYHDEKSPETKRAVKATISCEPAVFLASKIVRYVAVSMSRVOALNOMEDLOOSV
 PKYFLNTGCTDFPIKSNEMOALKMLNGRSMSEVPKPKETRMKYHEFVVDTL
 HLNRKRDPEPTNLTFGNATIVASRDVCHVLNPSQOLIEVKDTISFDEHLMA
 TLQARWMPGSVNPMPKIDYSDMTISLARLVKQWQEGDIDKGAFLAPCSGIHQRAICV
 YXAGDLMWMLQNHLLANFRDPKVDMDLQCLEYLRKAYIGTEL"
 BASE COUNT 359 a 307 c 322 g 329 t
 ORIGIN
 Query Match 14.1%; Score 191.8; DB 9; Length 1317;
 Best Local Similarity 52.5%; Pred. No. 3.6e-36;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;
 QY 317 tgaccagtgattgtgacattatcacagacttaagaggttaagtcaaaagctgtctcaa 376
 Db 317 TCACCAAGACTGTGACACTTCAAGGCTGAAGAGAACTTCAATACAGTCCCACTAGCA 376
 QY 377 aggaaggaagaaagcttcccaatagcctattcttggttgccaagaatgcaatatacg 436
 Db 377 AAGAAGAGGAGGAGTTCCTTATTCATCTATGATGATTCATGAGAGATGAAACT 436
 QY 437 ttgaagagcttaccatcagatataacaagacaataatttcacatcattatgac 496
 Db 437 TTGAAGGCTTACGCGAGCTGTGTATGCCCTAGAACATATCTGTGTCATGATGATG 496
 QY 497 gtaagacctgatacccttcaaaagtgcatagaacaatttagtaagtgtctccata 556
 Db 497 AGAAGTCCCGAAGAACTTCAAGAGCGGCTCAAGCAATATTCTGTGCTCCCAAG 556
 QY 557 ttctcattgttccaaatagaaagctgtgtaaatgcccacattccagactcaagctg 616
 Db 557 TCTTCATGCCAGTAACCTGTTCCGCTGCTTATGCTCTCTGCTCCAGGCTCAAGCTG 616
 QY 617 attaaattgctgtgcgaaccttcgaagcttcaatccagttggaatatgtatcaact 676
 Db 617 AACTCAACTGATGAGAGACTGTGCTCAGAGCTCAGTCCGCTGGAATACTCTGAATA 676
 QY 677 tgtgtggcaagatttccctcgaagcaaatlttgatgtgtcagaagtgaataaac 736
 Db 677 CATGTGGACGAGACTTTCCTATAAGAGCAATGACAGATGTGCCAGCTCTTAAGTGT 736
 QY 737 tcaatggaagcaaatatgttggagacggtgaaaccccccaacagtaaatgtgaaagattca 796
 Db 737 TGAATGGAGAAATAGCATGAGTCAAGAGTACCTCTTAACCAAGAAACCCGCTGGA 796
 QY 797 cttaaccatcagaacttgaacggtggtccttaataatgttaaacataaagacaa 856
 Db 797 AATATCACTTTGAGTGTAGAGACACATTAC-----CTAACCA 838
 QY 857 acatctccaaggaagaccocccataacatcagataattgttgcagtgctattttg 916
 Db 859 ACAAGAAGAGGATCCCTCCCTTATATTAATACATGTTTACAGGGATCCGACATG 898
 QY 917 tttaagtcagaacttgtttaaataatttcaacaacatccatcgttcaagaacttttg 976
 Db 899 TGGCTTCCCGAGATTTCCTCCAAACATGTTTGAAGAACCTTAATCCCACTGATG 958
 QY 977 cctgttcaaaagacacatactctcctgattgaagacttttgggttaacctgttggttc 1036
 Db 959 AATGGTAAAGACACTTAAGCCAGATGAACACTCTGGGGCCACCTCTACGCTGAC 1018
 QY 1037 caggaataacctggtggaatttccagatacagcccaagatg---tgtctgattcgaagta 1093


```

Db 923 AGAAGTCCCAAGAACTTCAAGAGGGGATCAAGCAATATTTCTTGTCCCAATG 982
Qy 557 ttcttgcctcccaatagagcgtggaatagccacattccagactcagagctg 616
Db 983 TCTTATAGCCAGTAAGTGCTGCGGTATATGCTCTGTCGCGGATGCAAGCTG 1042
Qy 617 atttaattgcttcgagcctctgaagcttcaatccagtggaataatttaact 676
Db 1043 ACCCTCAACTGATGAGAGCTTCTCCAGAGCTCAGTCCCTGGAAATCTTCCGAAATA 1102
Qy 677 tctgtgggcaagatttcccccgaagccaatttgaaattggtgcagagtggaataaac 736
Db 1103 CATGTGGAGGAGACTTTCTCTATAGAGCAATGACAGATGCTCCAGGCTTCAAGATGT 1162
Qy 737 tcaatggagcaaatatgttggagcgtggaaccccccaacagtaaatgtgaagatca 796
Db 1163 TGAATGGAGAGATGACATGAGTCAAGATACCTCTACAGCAAAAGAACCCGCTGGA 1222
Qy 797 cttaacatcatgaacttagacgggtgccttatgaatatgtgaagctaacataagaca 856
Db 1223 AATATCACTTTAGGTAGTAGAGAGACATATAC-----CTAACCA 1264
Qy 857 acatctccaaggaagaccccccaataacatcagatatgttggcagtgctattttg 916
Db 1265 ACAGAGAGAGATCCTCCCTTATATTTAACTATGTTTACAGGGAATGCGTACATTTG 1324
Qy 917 tttaagtaagcatttgttaatatatttcaaacactccgttgaagactttttg 976
Db 1325 TGGCTTCCCGAGATTTTCCCAACATGTTTGAAGAACCCCTTAATCCCAACACTGATTTG 1384
Qy 977 cctgtctaaagacacatactctcctgtatgagcatttggcctacattcggttc 1036
Db 1385 AATGGGTAAAGACACTTATACCCAGATGAACACCTCTGGGCGCACCTTCAGCTGCAC 1444
Qy 1037 caggaaatacctgggagatttccagatacagccagagtg---tgcitacatgcagata 1093
Db 1445 GGTGATGCTTGGCTCTGTCCCAACACACCCCAAGTACAGACTCTCAACATGACTTTTA 1504
Qy 1094 agactgcctgtcaagtggaattactatgaagcctttctatccagt----- 1143
Db 1505 TTGCCAGGCTGTCAAGTGGCAGGGTCAATGAGGAGACATCGATTAAGGCTCCTTATG 1564
Qy 1144 -----tgacttgatcttcacctgaagcgtgtatatttagagacgtcagaataggt 1198
Db 1565 CTCCTGCTCTGGAATCCACGCGGCTATCTGCGTTATGGGCTGGGACTTGGAAT 1624
Qy 1199 ggcctatcaaatgagacattgttgcctaataatttgatcttaaggctgaacctatct 1258
Db 1625 GGATGCTTCAAAACCATCACCTGTTGGCCACAGATTGACCCAAAGGTAGATGATG 1684
Qy 1259 tgattaaatgcttgcagaaagct 1283
Db 1685 CTCTTCACTGCTTAGAAGAAATACCT 1709

```

RESULT 12
 AC022480/C
 ID AC022480 standard: DNA: HTG: 181983 BP.
 XX
 AC AC022480;
 XX
 SV AC022480.5
 XX
 XX
 DT 06-FEB-2000 (Rel. 62, Created)
 DT 18-JUL-2000 (Rel. 64, Last updated, Version 5)
 XX
 DE Homo sapiens chromosome 15 clone RP11-361D15, WORKING DRAFT SEQUENCE, 24
 DE unordered pieces.
 XX
 KW HTG: HTGS_DRAFT; HTGS_PHASE1.
 XX
 OS Homo sapiens (human)

```

OC Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;  

OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  

XX  

XX [1]  

RP 1-181983  

RA Waterston R.H.;  

RT "The sequence of Homo sapiens clone";  

RL Unpublished.  

XX  

XX [2]  

RP 1-181983  

RA Waterston R.H.;  

RT Submitted (04-FEB-2000) to the EMBL/GenBank/DBJ databases.  

RL Genome Sequencing Center, Washington University School of Medicine, 4444  

RL Forest Park Parkway, St. Louis, MO 63108, USA  

XX  

XX On Jul 16, 2000 this sequence version replaced gi:7321974.  

CC  

CC Genome Center  

CC Center: Washington University Genome Sequencing Center  

CC Center code: WUGSC  

CC Web site: http://genome.wustl.edu/gsc/index.shtml  

CC  

CC Project Information  

CC Center project name: H_NH0361D15  

CC  

CC Sequencing vector: M13; 100%  

CC  

CC Sequencing method: Dye-terminator Big Dye; 0% of reads  

CC Chemistry: Dye-terminator Big Dye; 0% of reads  

CC Assembly program: Phrap; version 0.990319  

CC Consensus quality: 164956 bases at least Q40  

CC Consensus quality: 170052 bases at least Q30  

CC Consensus quality: 172655 bases at least Q20  

CC Insert size: 168000; agarose-fp  

CC  

CC Quality coverage: 3.76 in Q20 bases; agarose-fp  

CC Quality coverage: 3.85 in Q20 bases; sum-of-coverage  

CC  

CC  

CC * NOTE: This is a 'working draft' sequence. It currently  

CC * consists of 24 contigs. The true order of the pieces  

CC * is not known and their order in this sequence record is  

CC * arbitrary. Gaps between the contigs are represented as  

CC * runs of N, but the exact sizes of the gaps are unknown.  

CC * This record will be updated with the finished sequence  

CC * as soon as it is available and the accession number will  

CC * be preserved.  

CC  

CC 1  

CC 1087: contig of 1087 bp in length  

CC *  

CC 1088  

CC 1187: gap of unknown length  

CC *  

CC 1188  

CC 2741: contig of 1554 bp in length  

CC *  

CC 2742  

CC 2841: gap of unknown length  

CC *  

CC 2842  

CC 4520: contig of 1679 bp in length  

CC *  

CC 4521  

CC 4620: gap of unknown length  

CC *  

CC 4621  

CC 6998: contig of 2378 bp in length  

CC *  

CC 6999  

CC 7096: gap of unknown length  

CC *  

CC 7097  

CC 10122: contig of 3024 bp in length  

CC *  

CC 10123  

CC 10222: gap of unknown length  

CC *  

CC 10223  

CC 14393: contig of 4171 bp in length  

CC *  

CC 14394  

CC 14493: gap of unknown length  

CC *  

CC 14494  

CC 18603: contig of 4110 bp in length  

CC *  

CC 18604  

CC 18703: gap of unknown length  

CC *  

CC 18704  

CC 22356: contig of 3653 bp in length  

CC *  

CC 22357  

CC 22456: gap of unknown length  

CC *  

CC 22457  

CC 26948: contig of 4492 bp in length  

CC *  

CC 26949  

CC 27048: gap of unknown length  

CC *  

CC 27049  

CC 31773: contig of 4725 bp in length  

CC *  

CC 31774  

CC 31873: gap of unknown length  

CC *  

CC 31874  

CC 36373: contig of 4500 bp in length  

CC *  

CC 36374  

CC 36473: gap of unknown length  

CC *  

CC 36474  

CC 36474: gap of 5089 bp in length  

CC *  

CC 41563  

CC 41562: gap of unknown length  

CC *  

CC 41563  

CC 46645: contig of 4983 bp in length  

CC *  

CC 46646  

CC 46745: gap of unknown length  

CC *  

CC 46746  

CC 53335: contig of 6590 bp in length

```

FT		/note="assembly_name:Contig45 clone_end:SP6
FT		vector_side:left
FT	misc_feature	159783..181983
FT		/note="assembly_name:Contig46"
xx		
SQ	Sequence 181983 BP; 53047 A; 36975 C; 36896 G; 52718 T; 2347 other:	
	Query Match	14.1%; Score 191.8; DB 33; Length 181983;
	Best Local Similarity	52.5%; Pred. No. 4.3e-36;
	Matches 517; Conservative	0; Mismatches 432; Indels 36; Gaps 3
QY	317 taccacgattgttgaacttatcagaccttaaggagtatgcctcaaaagcttgctcaa	376
Db	86723 TCACCAAGACTGTGAGCCTTCAAGGCTGAAGAAGATTCAATACGTTCCACATGACA	86664
QY	377 aggagagaanaagctccccaatagcctatctttgtygtgccaaaagatgaattag	436
Db	86663 AAGAAGAGTGAGATTCCTTATGGCATCTATGCGTAGTCATGAGAGATTGAATAA	86604
QY	437 ttgaaaggtttccatgctatatataaacacagacaattttacggatcccatatgac	496
Db	86603 TTGAAGGCTACTGCAGGCTGTGATGCCCTCAGAACATATACGTGTGCCATGTGGATG	86544
QY	497 gtaagcacctgatacccttcaaatgcttcgatgaacaatttagctaagttcctccoa	556
Db	86543 AGAAGTCCCCAGAACTTTCAAAGAGGGCGTCAAAAGCANTTTATCTTGCTCCC	86484
QY	557 tttcatagtcttccaaattaagagcgctgygaatatyccccacattccagactccagctg	616
Db	86483 TCTTCATACCAGTAAGCTGTGGGTGGGTGTTATGCTTCCTCGTCCAGGATGCAAGCTG	86424
QY	617 attaatgtctgtcgagcccttcygaagcttccaatccagtcagtggaatatgttacaact	676
Db	86423 ACCTCACTGCATGAGAAGCTTGCTCCAGAGCTCACTGCCGTGGAATACTTCCIGATA	86364
QY	677 tgtytgagcaagaatttcccctgaagtcaaatlttgaatttggtygtcagagttgaanaac	736
Db	86363 CATGTGGAGCGGCTTCTCTTAATAAAGCAATGACAGATGTGGCCAGCTCCAGATGT	86304
QY	737 tcaatgagcaaatatagtgttgagagcgygtiaaaccocccaaacagtaaattggaagtta	796
Db	86303 TGAAATGGAGGATATACATGAGTCAAGAGTACCTCTTAAGCAACAAGAACCCGTGGA	86244
QY	797 ctacacataatgaacttgaacggygttccttatgaatatytgaaagtcacaataagaca	856
Db	86243 AATATCACTTTGAGTAGTAGAGACACATTTACAC-----CTAACCA	86202
QY	857 aaatcccaaggaagacccccccataacattcaagtatgttggcagtgcttatttg	916
Db	86201 ACAAGAAGAGGATCTCTCCCTCTTAATAATTAATACTATGTTTACAGGAAATGCTTACATTTG	86142
QY	917 tttaaagtaagcatttgtttaatatattttcaacaacccacatcgcttcaagaacttttg	976
Db	86141 TGGCTTCCGAGATTTGTCACCAATGTTTGAAGAACCCTTAATCCCAACACATGATGG	86082
QY	977 cctgtgcttaagacaatatctctcctgtagagagacttttgggttaacttgatcgggttc	1036
Db	86081 AATGGGTAAAAAGCACTTTATAGCCCAAGATGAACCTCTGTGGCGACCTTCACGCTGACAC	86022
QY	1037 caagaatacctggggagatttccagatacagcccaagatg---tgctgaictgcagaagta	1093
Db	86021 GGATGATGCTGTGCTGTGTTCCCAACACCCCAAGNACGACATCTCAGACATCTTCTTA	85962
QY	1094 agaactgccttgcagaatggaattactatgaaggtttttctatccagt-----	1143
Db	85961 TTGCCAGGGCTGGCAAGTGGCAAGGCTCAATGAGGAGACATCGTAAGGGGCTCCTTATG	85902
QY	1144 ----tgtaactgatactacccttcygaagcgttgytatattatgagcttgcaaataggt	1198
Db	85901 CTCCTGCTCTGGAATCCACACAGCGGCTATCTGCGTTATGGGCGCTGGGACTTGAAT	85842
QY	1199 ggcttaccaagaatggacatttgglttgctataataaatttgattcctaagtggnacotactct	1258

```

Db 85841 GGAGCTTCAAAACCATCCTGCTGGCCAAAGATTGACCAAGATGATGATATG 85782
QY 1259 tgattaatgcttgcgcagaaagct 1283
Db 85781 CTCTCAGTCTGTAGAAATACCT 85757

RESULT 13
AC092755
LOCUS AC092755
DEFINITION Homo sapiens chromosome 15 clone RP11-361D15 map 15q22, ***
ACCESSION AC092755 AC022480
VERSION AC092755.2 GI:15022677
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 187760)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.
Pate, D. and Hood, L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 187760)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.
Pate, D. and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT On Jul 27, 2001 this sequence version replaced g1:15021993.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMSC
Web site: http://chroma.mbl.washington.edu/msg_xww
Contact: leerowens@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Assembly program: Phrap; version 0.990399
Insert size: 185000; agarose-efp
Quality coverage: 10.4x in Q20 bases; sum-of-contigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 6055: contig of 6055 bp in length
* 6056 6155: gap of unknown length
* 6156 33287: contig of 27132 bp in length
* 33288 33388: gap of unknown length
* 33388 160326: contig of 126839 bp in length
* 160326 160327: gap of unknown length

```

```

FEATURES
* 160327 187760: contig of 27434 bp in length.
Source
1. 187760
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q22"
/clone="RP11-361D15"
/clone_lib="RP11-361D15"
/note="This clone overlaps RP11-112N19 and RP11-568G20"
BASE COUNT 53775 a 38839 c 38241 g 56604 t 301 others
ORIGIN
Query Match 14.1% Score 191.8 DB 2: Length 187760;
Best Local Similarity 52.5% Pred No. 4.3e-36;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;
QY 317 tgaccagtgatgtgacattatccagactctaagaggtatgtcgaagctgtctcaa 376
Db 40861 TCACGACAGACTGTGAGACTTCAAGGCTGAAGAGAACTCATACAGTCCACAGCA 40920
QY 377 agagagaagaagcttcccaatagctatcttgggtgtccacaagaatgaattagc 436
Db 40921 AAGAAAGGTGAGATTCCTATTGCTACTGATGATGATGATGATGATGATGATG 40980
QY 437 ttgaagagcttaccatataccacacacacacacacacacacacacacacacac 496
Db 40981 TTGAAGGCTACTGACGAGTGTATGATGATGATGATGATGATGATGATGATG 41040
QY 497 gtaagagacactgacattcccaagtgtccatgaacacacacacacacacacac 556
Db 41041 AAGAGTCCCGAGAACTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 41100
QY 557 ttctcctgtcccaaatagagagcgtgtgaatgatccacacacacacacacac 616
Db 41101 TCTTCATACGACATGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 41160
QY 617 attaaattgctgtgtgagaccttcgaagcttcacacacacacacacacacac 676
Db 41161 ACCTCACTGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 41220
QY 677 tggtaggaagagcttcccccgaagacacacacacacacacacacacacacacac 736
Db 41221 CATGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 41280
QY 737 tcaatgagacaaatattgttggagagcgtgaacacacacacacacacacac 796
Db 41281 TGAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 41340
QY 797 ctgacacatagacttagagagcgtgtccttgaatgtgagacacacacacacac 856
Db 41341 AATATACCTTTGAGTAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 41382
QY 857 acatctcccaagagagacacacacacacacacacacacacacacacacacac 916
Db 41383 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 41442
QY 917 tttaagtaagatttgttaataatatttcaaacacacacacacacacacac 976
Db 41443 TGGCTTCGAGAGATTTGCTGACACATGTTTGAAGAGACCTTAATCCCAACATG 41502
QY 977 ccgtgacaaagagacacacacacacacacacacacacacacacacacacac 1036
Db 41503 AATGGGTAAAGAGACTTATAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 41562
QY 1037 caggaataactgtggagagatttccagatccacagagatgtgtgtatcgcag 1093
Db 41563 GGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 41622
QY 1094 agactgaccttcaagatgaattactatgaagagcttttccatcccaatgtt 1143
Db 41623 TTGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 41682

```

QY	1144	--	ttctactaggatctccactcttgaaacgcgtgatattatcaggagctccaaattaaagt	1198
Db	41683	CTCCCTGCTCTGGAAATCCACACGCGGGCTATCTGCGTTTAAGGGGCTGGGACCTTGAATT	41742	
QY	1199	gacctataaagaatggacattggtttgctataaattgattcttaagtggaacctatct	1258	
Db	41743	GGAAAGCTCAAAACCATCACCCTGTGGCCACCAAGTTTGACCCCAAGAGTAGATGATATG	41802	
QY	1259	tgattaatgcttgccagaanaagct	1283	
Db	41803	CTCTCACTGCTTAGAAGAAATACCT	41827	
RESULT	14			
AF231105			VRL	26-MAY-2000
LOCUS				
DEFINITION	AF231105	2017 bp	DNA	
ACCESSION	AF231105			
VERSION	AF231105.1	GI:8096688		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				

	Matches	518:	Conservative	0:	Mismatches	429	Indels	39:	Gaps	3
QY	316	atgacacagatgattgtagacttattcaagctcttaagaggttatgctcaaaagcttgctca	375							
Db	551	ATATACCAAGGAGACTGTGAGCGCTTCAAGGCCCAAAAGAAATGATATACAGTTCACATGAGT	610							
QY	376	aagagagagaaaaagctctcccaataagaccattctcttggttgctccaaagaatgcaattatg	435							
Db	611	AAAGAAGAGGTGAGACTTCCCTCCATTCCTTACCGATGGTGCTCATGAGAAGATTGGAAC	670							
QY	436	gttgaaagcttatccatgctctatatacaaaccaagcaaatattactgcatccattatgat	495							
Db	671	TTTGAACGGCTGCTGGAGCCGCTGATGCCCTCGAAGCATATATCTGTCTCACGTGGAT	730							
QY	496	cgtaaagcaccctgataccttcaaaagtgcataatgcaaatattgaatgctcttccaat	555							
Db	731	GTAAGATCCCCAGACACTTTCAAAAGCGGGTCAAGGCCATTATTTCTGCTTCCCAAT	790							
QY	556	attctatgcttcccaaatlaagagctgtgaaatagatgcccaatttcccaagcccaaggt	615							
Db	791	GCTTTATATGGCCACATPAGTTGTTCCCGGTGGTTTATGCTCTCTGTGCTCAAGATGCAAGCT	850							
QY	616	gatttaaatgtctgtgcgacccttcgaagcttcaatccaaatgcaatgtagaaatgattatcaac	675							
Db	851	GACCTTAACCTGATGTGAAGACTTGCCTCCAGAGCTCAGAGTCATGATGAGTACTTACGAAAT	910							
QY	676	tgtgtgtggcgaagatttccctgaagcgaatttgatattgattgtgtcagatgtgaaaaaa	735							
Db	911	ACATGGGGACACACTTCCCATTAAGACCAATGCCAGATGGTCTGGCCCTCAAGATG	970							
QY	736	ctcaatgysgcaaatatgtttgagaacggtlgaaccccccaagtaaatgtgsaaagatc	795							
Db	971	TTTGAAGGGTAAAGACAGTATGAGAGTCTGAGAGACTTCTGATGTCACAAAATAATTCCTGG	1030							
QY	796	acttaccatcaatgaacttgaagcaggttgcttatgatatgtgaatgysgccaataagaca	855							
Db	1031	AAATACCCTATAGAGTGTGACACACACTGATACCTTACACCAAGATA-----	1078							
QY	856	aacatctccaagaagacaccccccaataaacttaagatattgttgtagtcttaattt	915							
Db	1079	-----AAAGACCCCTCCCTCCCTATATTATTTACCCATGTTCAACAGGGAATGCCCTATT	1129							
QY	916	gttttaagtcgaagcattgtttaatatatttcaacaactcatcagctcgaagctttt	975							
Db	1130	GTTGGCTCTCGAACCCTTTTCCAAATGTCTTAGACAACCTTAATCCCAATATCTGATT	1189							
QY	976	gacctgtctaagaacacatactcttcgtatgagcaatttvggctacattcggtt	1035							
Db	1190	GAATGGGTCAAAAMCACATATAGCCCCACAGACACACCTCTGGGACACCTTCAGCGTCT	1249							
QY	1036	caaagaaataccgtgggaagattccagatcagcccaagatg--tgcctgatactgcagagt	1092							
Db	1250	CCGTGATGCTTGGCTGTCTTCTTCCAGCACCCCAAGATATCAATCTGACATGAGTACGTGC	1309							
QY	1093	aagaatcgcgcctgtcaagtgtgaattactatataaggtctttc-----tat	1137							
Db	1310	ATGCGCAGGCTGCTCAAGTGGCAGTACACAGAGGAGATGTGCAGATGGGGCGCGCTTAT	1369							
QY	1138	cccaagtctgaactgacatcacccttcgaagcgtgtatatttagagctgtcagaattaag	1197							
Db	1370	GCACCTGCTCTGAAATCATTCATGGAGGGCCATCTGCATTATACGGGGCGGGGAACTGTAC	1429							
QY	1198	tgtgcttatcaagaatgagacatgtgtttgtctaataatttgatcttcaagtgtgaacctatc	1257							
Db	1430	TGGATTCTCCAGAACCATATACCTCTTGGCAAAACAAGTTTGACCCGAGGAGTGAATGATAAC	1489							
QY	1258	ttagtaaatgcttgcgcagaanaagct	1283							
Db	1490	GTCCTGAGTGTCTTACGAAGATACCT	1515							

RESULT

15

RM058182

Fri May 3 10:57:51 2002

us-09-645-192-1.rge

Page 17

Page 1

OM nucleic - nucleic search, using SW model

Run on: May 2, 2002, 09:33:00 ; Search time 175.33 Seconds

6659.881 Million cell updates/sec

Title: US-09-645-192-1

Sequence: 1 atgaagatatccaatgtta.....atctcactaccaacatcatga 1362

Scoring table: IDENTITY_NUC

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Maximum DB seq length: 200000000000

Post-processing: Minimum match 0%

Listing first 45 summaries

Database :

1:	/SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT *
2:	/SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT *
3:	/SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT *
4:	/SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT *
5:	/SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT *
6:	/SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT *
7:	/SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT *
8:	/SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT *
9:	/SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT *
10:	/SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT *
11:	/SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT *
12:	/SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT *
13:	/SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT *
14:	/SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT *
15:	/SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT *
16:	/SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT *
17:	/SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT *
18:	/SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT *
19:	/SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT *
20:	/SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT *
21:	/SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT *
22:	/SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1362	100.0	1362	22	AA800045	Human DNA encoding
2	503	36.9	781	22	AA98678	Human EST-derived
3	191.8	14.1	1221	21	AAA96571	Core 2 beta-1,6-N
4	191.8	14.1	1317	21	AAA96569	DNA encoding a
5	191.8	14.1	2108	21	AAA96570	Core 2 beta-1,6
6	191.8	14.1	2128	22	AAC62134	Nucleotide sequen
7	191.8	14.1	2229	21	AA949109	Human pancreatic c
8	191.8	14.1	2235	22	AAH34453	Human colon cancer
9	191.8	14.1	2319	21	AAA46623	Human C2/4GRT CDNA
10	185.2	13.6	1807	21	AA083201	l-branched enzym
11	185.2	13.6	1807	17	AA716201	Beta-1,6-N-acetylgl

12	185.2	13.6	1807	19	AAV30006	Full length CDNA
13	185.2	13.6	1807	19	AAV130006	Human beta-1,6-N-ac
14	183	13.4	186	21	AACT9209	Human secreted protein
15	175.2	12.9	2105	15	AAQ61559	CDNA sequence of H
16	174.4	12.8	5010	15	AAZ4042	Rat DHI CDNA. Rat
17	174.4	12.8	5010	21	AAZ65468	Diabetic rat heart
18	172	12.6	2204	19	AACT2403	Human core 2Gnt DR
19	172	12.6	2204	21	AACT65469	Human heart core 2
20	160.8	11.8	2102	18	AACT80112	Coding sequence f
21	121.2	8.9	378	19	AAV30005	CDNA encoding a r
22	112.4	8.3	936	22	AAV58252	Oligonucleotide D
23	112.4	8.3	936	22	AAV58254	Oligonucleotide D
24	112.4	8.3	936	22	AAV58257	Oligonucleotide D
25	112.4	8.3	936	22	AAV58259	Oligonucleotide D
26	112.4	8.3	936	22	AAV58262	Oligonucleotide D
27	112.4	8.3	936	22	AAV58255	Oligonucleotide D
28	110.6	8.1	936	22	AAV58254	Oligonucleotide D
29	110.6	8.1	936	22	AAV58252	Oligonucleotide D
30	110.6	8.1	936	22	AAV58257	Oligonucleotide D
31	110.6	8.1	936	22	AAV58259	Oligonucleotide D
32	110.6	8.1	936	22	AAV58265	Oligonucleotide D
33	110.6	8.1	936	22	AAV58255	Oligonucleotide D
34	92.2	6.8	378	17	AAAT16202	C2Gnt catalytic di
35	86.8	5.9	997	19	AAV59800	Human secreted p
36	86.2	4.9	300	21	AAA01261	Human colon cancer
37	53.4	3.6	192	22	AAAC6139	Murine beta-1-6-N
38	49.4	3.6	777	19	AAV56681	Human secreted p
39	47.4	3.5	307	21	AAAC01854	Streptococcus pneu
40	46.2	3.4	6171	19	AAV57170	Oligonucleotide D
41	41.8	3.1	244	22	AAV58238	Oligonucleotide D
42	40.6	3.0	244	22	AAV58238	Oligonucleotide D
43	40.6	3.0	7430	19	AAV31250	E. coli y96 pathog
44	36.8	2.7	466	22	AAAT18302	Probe #6988 used
45	35.8	2.6	623	22	AAAT12685	Human CDNA clone

ALIGNMENTS

RESULT

ID AAS00045 standard; DNA; 1362 BP.

AC AAS00045;

DT 14-MAY-2001 (first entry)

Human DNA encoding C2GnT3.

Human; C26nt3; Thymus-related disorder; cancer; tumour; adenoma;

KW malignant melanoma; breast cancer; cervical cancer; hypoadrenalism;
KW hyperactivity; atrophy; thymus gland; autoimmune; antibody; disease;

acquired immunodeficiency syndrome; sepsis; wound healing; infection; de
leukemia; lymphoma; AIDS; Wiskott-Aldrich syndrome;
KW
KW

Homo sapiens

Key	Location/Qualifiers
AA	
EH	

$\frac{1}{\star t g q} =$

primer_bind	/product=	220113
	complement	(114..138)

FT 2 3
/label= "Binding site for PCR primer TSHC100"

```

-----
FT
/*tag= c

```

XX

XX

XX

PF 24-AUG-2000; 2000MO-DK00469.
XX
PR 24-AUG-1999; 99US-0150488.
XX
PA (SCHW/) SCHWIENTEK T.
PA (CLAU/) CLAUSEN H.
XX
XX
PI Schwiientek T, Clausen H;
XX
XX WPI; 2001-226615/23.
DR P-PSDB; AAU00037.
DR
XX
XX
PT New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
PT thymus-related disorders, cancers, tumours, immunosuppression -
XX
XX Claim 5; Fig 1; 97P; English.
DS

The sequence encodes Human UDP-N-acetyl-glucosylamine-6-phosphate-3-N-acetylglactosamine-alpha-R-beta1-6-N-acetylglucosaminyltransferase (UDP-GlcNAc 6Galbeta1,3GlcNAc alpha-R betaal,6GlcNAc transferase or C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the preparation of compositions for treating a conditions mediated by C2GnT3, particularly a thymus-related disorder. C2GnT3, nucleic acids encoding it and antibodies against it may also be used for in vitro purposes related to scientific research, DNA synthesis and manufacture of vectors, in the prognostic and diagnostic evaluation of conditions associated with altered expression or activity of C2GnT3 or conditions requiring modulation of C2GnT3, as well as in monitoring conditions by detecting and localizing the DNA and protein. Disorders such as tumors (e.g., malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the breast or cervix), hyporeactivity, hyperactivity, atrophy, enlargement of thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression, acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome, sepsis, wound healing, acute and chronic infection, cell-mediated or humoral immunity, or TH1/TH2 imbalance, may be treated using these protein or nucleic acid. The antibodies may be used to screen potential therapeutic compounds to determine their effects on conditions such as thymus-related disorder or cancer, to determine the level of C2GnT3 expression in cells genetically engineered to produce C2GnT3, or to detect and quantify polypeptides in a sample to determine their role in a particular cellular events or pathological states and to diagnose and treat such pathological states.

Sequence 1362 BP; 410 A; 260 C; 279 G; 413 T; 0 other;

Query Match	100.0%	Score 1362;	DB 22;	Length 1362;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1362; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

QY	1	atgaagaatttcaaaatgttatctttaaaccattccctcaagcagaagaatttcaaccttt	60
Db	1	atgaagaatttcaaaatgttatctttaaaccattccctcaagcagaagaatttcaaccttt	60
QY	61	ttaacctatgctgctctctctttgtttaaagcttcaatgvtgagcagcactcttcgcgaa	120
Db	61	ttaacctatgctgctctctctttgtttaaagcttcaatgvtgagcagcactcttcgcgaa	120
QY	121	aagacattacttggtgtgagtaacctccctaagctacccctgcctttgttaagaacaagatac	180
Db	121	aagacattacttggtgtgagtaacctccctaagctacccctgcctttgttaagaacaagatac	180
QY	181	actcagctttaaagatgagatgcagatcagatcgaagatgaactgtgtcgggattatcagaacagag	240
Db	181	actcagctttaaagatgagatgcagatcagatcgaagatgaactgtgtcgggattatcagaacagag	240
QY	241	cctttggaattgggaagaagctctcggaataagaagaaggacatcattgactttggaagat	300
Db	241	cctttggaattgggaagaagctctcggaataagaagaaggacatcattgactttggaagat	300
QY	301	gagatgctgtgtggcaatgacagatgattgtgacatttatcagactcaagcgttgaatgct	360
Db	301	gagatgctgtgtggcaatgacagatgattgtgacatttatcagactcaagcgttgaatgct	360

Db	301	gaigatgtgtgccaatgaccagtgatttctgtgcatttacagacccctaaagagttatgct	360
Qy	361	caaaagctgtctccaagagggagaaagctctcccaatgacctattcttggttgccac	420
Db	361	caaaagctgtctccaagagggagaaagctctcccaatgacctattcttggttgccac	420
Qy	421	aaagatgcacattatgtgtgaagaagctctacacatgcatataaacacagacacatatcc	480
Db	421	aaagatgcacattatgtgtgaagaagctctacacatgcatataaacacagacacatatcc	480
Qy	481	tgcgtccattatgatgcgttaagggacctgtataccttcaaaagttgcacatgaacaattagct	540
Db	481	tgcgtccattatgatgcgttaagggacctgtataccttcaaaagttgcacatgaacaattagct	540
Qy	541	aaggtgctctcccaatattttccattgtcttcocaaattgaagagctgtgtgaattgcacatt	600
Db	541	aaggtgctctcccaatattttccattgtcttcocaaattgaagagctgtgtgaattgcacatt	600
Qy	601	tccgacctccaggctgatttaatttctgtgtgcagacctctgaattcttaattccaacaggg	660
Db	601	tccgacctccaggctgatttaatttctgtgtgcagacctctgaattcttaattccaacaggg	660
Qy	661	aaatagtctacaaactgtgtgtggccaagaatttcacctgaagtcacaatttgaattgtgtg	720
Db	661	aaatagtctacaaactgtgtgtggccaagaatttcacctgaagtcacaatttgaattgtgtg	720
Qy	721	tcaagcttgtaaaaaaacctcaatgtgagaacaatagtgtgagaaggttaaaoccccaaacgt	780
Db	721	tcaagcttgtaaaaaaacctcaatgtgagaacaatagtgtgagaaggttaaaoccccaaacgt	780
Qy	781	aaattggaagagatcacttcaccaatcgaactgaacgggtgccttatgaatattgtgaag	840
Db	781	aaattggaagagatcacttcaccaatcgaactgaacgggtgccttatgaatattgtgaag	840
Qy	841	ctaccacaaaggacaaacacatcccaagggagaccccccccaataaacatccagattttgt	900
Db	841	ctaccacaaaggacaaacacatcccaagggagaccccccccaataaacatccagattttgt	900
Qy	901	ggacgagctattttgttttaagtccaagaattgtttaaatatattttcaacaactccatc	960
Db	901	ggacgagctattttgttttaagtccaagaattgtttaaatatattttcaacaactccatc	960
Qy	961	gttcaagaacttttttgcctgtgtctaaagacacatctccgtagagagacatttgggct	1020
Db	961	gttcaagaacttttttgcctgtgtctaaagacacatctccgtagagagacatttgggct	1020
Qy	1021	accttgatccgggttccagaataacatggggagatttccagatcagcccgagatgtgct	1080
Db	1021	accttgatccgggttccagaataacatggggagatttccagatcagcccgagatgtgct	1080
Qy	1081	gactgtgaagagatgaagactgcctctgcacaagtgaatattacatgaagagctttttccatcc	1140
Db	1081	gactgtgaagagatgaagactgcctctgcacaagtgaatattacatgaagagctttttccatcc	1140
Qy	1141	agttgtacgcgcatctcaaccttcogaaagcggtgtgatttalttgaggtcgcagaaatlaaggtg	1200
Db	1141	agttgtacgcgcatctcaaccttcogaaagcggtgtgatttalttgaggtcgcagaaatlaaggtg	1200
Qy	1201	ctttcaaaagatgcacattggttctgttaataattgattcttcaaggttgagccctatctgt	1260
Db	1201	ctttcaaaagatgcacattggttctgttaataattgattcttcaaggttgagccctatctgt	1260
Qy	1261	attaaatgcttgcagaaagaagcttgaagaacagcagagagactgatacatttgccctca	1320
Db	1261	attaaatgcttgcagaaagaagcttgaagaacagcagagagactgatacatttgccctca	1320
Qy	1321	gaaagattattatgtgataagaataatccatatacacatatga	1362
Db	1321	gaaagattattatgtgataagaataatccatatacacatatga	1362
RESULT	2		
FAH98678			

RESULT
AAH98678

ID		AAH98678 standard; cDNA; 781 BP.
AC		AAH98678;
DJ		12-Oct-2001 (first entry)
DE		Human EST-derived coding sequence SEQ ID NO: 535.
KM		Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse; tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW		diagnostics; forensic test; gene mapping; genetic disorder;
KX		biodiversity; gene therapy; nutrition; ss.
OS		Homo sapiens.
XX		WO200154477-A2.
PD		02-Aug-2001.
PF		25-JAN-2001; 2001MO-USO2687.
PR		25-JAN-2000; 2000US-0491404. 17-Jul-2000; 2000US-0617746. 03-AUG-2000; 2000US-0631451. 15-SEP-2000; 2000US-0663870.
PA	(HYSE-) HYSEQ INC.	
PI	Tang YT, Liu C, Zhou P, Qian XE, Wang Z, Chen R, Asundi V;	
FI	Cao Y, Drmanac RA, Zhang J, Werhman T;	
DR	WP1; 2001-476164/51.	
PS	P-PDSB; AAM24019.	
PT	Isolated polypeptide for treatment of diseases, diagnostics, raising antibodies and research use -	
CC	Claim 1; Page 549; 1275pp; English.	
XX	The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a cDNA of the invention.	
SQ	Sequence 781 BP; 218 A; 175 G; 174 C; 214 T; 0 other;	
Query Match	36.9%; Score 503; DB 22; Length 781;	
Best Local Similarity	96.3%; Pred No. 1,4e+131;	
Matches 515; Conservative 0; Mismatches 20; Indels 0; Gaps 0		
OY	1 atgaagatattcaaatgttatctttaaacatacctaaccagaagaagtttcatccgittt 60	
Db	230 atgaagatatccaaitgttatctttaaacatacctaaccagaagaagtlttcactcgitt 289	
OY	61 ttaaccttaggcggctcctcttgtaaaggcttcaaagtatagacgactttccgaa 120	
Db	290 ttacccttaggcgccctcttgttaaaggcttcaaagtatagacgactttccgaa 349	
OY	121 aaagacaattaactgtgttgagtaccccctaagtaactcgcctttgttaagaacaagat ac 180	
Db	350 aaaagacattactcgtgttgagtalccccctaagtaactcgcctttgttaagaacaagat ac 409	
OY	181 actatgttaagatgaagtcagatgaagttaactgttcgggatcatcgaacaag ag 240	
Db	410 acctatgttaagatgaagtcagatgaagttaactgttcgggatcatcgaacaag ag 469	
OY	241 ccttgaagaattggaagaagctcgaataagaagaaggacatcattgacttggaagat 300	

Db 470 cctttagaatttgaaagagctgtctgaataaagaagaaggagccatcattgcttggagat 529

Qy 301 gatgattcttggcaatgacccagtgattgttgacattctcagactcctaagaagttatgct 360

Db 530 gatgattcttggcaatgacccagtgattgttgacattctcagactcctaagaagttatgct 589

Qy 361 caaagcttctcccaagaagagagaaaagctctccccaagaactcttcttggttgtccac 420

Db 590 taagaagcttcctcaagaagagagaaaacctctccccaagaactcttcttggttgtccac 649

Qy 421 aaagatgcattatgcttgaagagcttatccatgtctatatacaaccagcaaatattac 480

Db 650 caaagaacatcatggtgtgagagcttatccatgtctatatacaaccagcaaatattac 709

Db 710 tgcaccattatgagcgaggagcactgaaccttcaagttgcctgaacoatt 764

RESULT 3

AA96575

ID AA96575 standard; DNA; 1221 BP.

AC

XX AAA96575;

DT 08-FEB-2001 (first entry)

XX

DE Core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region DNM.

XX

XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;

KM cancer; cardiovascular disorder; inflammatory disorder; asthma;

KM rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;

KM septic shock; adult respiratory distress syndrome; ARDS; cancer;

KM platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;

KM clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;

KM diverticulitis; ulcerative colitis; SS.

XX

OS Homo sapiens.

XX

PN CA2296936-A1.

XX

PD 03-AUG-2000.

XX

PF 03-FEB-2000; 2000CA-2296936.

XX

PR 03-FEB-1999; 99US-0118674.

XX

PA (GLYC-) GLYCDESIGN INC.

XX

PI Korczak B, Lew A;

XX

DR WPI; 2000-594746/57.

XX

PT New nucleic acid molecules of core 2

PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new

PT compositions for treatment of disorders mediated by the enzyme

PT including cancer, cardiovascular and inflammatory disorders.

XX

PS Claim 4; Page 54; 66pp; English.

XX

XX The present sequence encodes a fragment of a human core 2

CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide

CC The polypeptide can be used to treat diseases and disorders, such as

CC cancer, cardiovascular disorders and inflammatory disorders including

CC asthma, rheumatoid arthritis, inflammatory bowel disease,

CC arteriosclerosis, septic shock, adult respiratory distress syndrome

CC (ARDS) and cancer. Various platelet-mediated pathologies such as

CC atherosclerosis and clotting can also be treated. The polypeptides of

CC the invention are predominantly expressed in gastrointestinal tissue

CC (stomach, colon, intestine, testis) and are elevated in cancer.

CC Gastrointestinal disorders that may be prevented or treated include

CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis

CC and ulcerative colitis. The antibodies may be used in

CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.

XX
 SQ Sequence 1221 BP; 342 A; 285 C; 295 G; 299 T; 0 other;

Query Match 14.1%; Score 191.8; DB 21; Length 1221;
 Best Local Similarity 52.5%; Pred. No. 5,7e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

QY 317 tgaccagtgttggacattatcagacttaagaagttatgctccaanaactgtgtccaa 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221 tcaccagagactgtgagacttcaaggctcaagaagttacatacagctccactgagca 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 377 agagaggaanaagcttcccaataagcctatcttgggtgtgccaaagatgcaattatg 436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 281 aagaagaggttgagttccctatgtccatctcatgtatgcatcagagaattgaaact 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 437 ttgaagagctatccatgcatataacaacagacacataattctgcatccattatgac 436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 341 ttgaagagctatccatgcatataacaacagacacataattctgcatccattatgac 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 497 gtaaggaactgtacaccttcaaaagtgcacagaaattagcaatgagtgcttcccaata 556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 401 agaagtcocccagaacttcccaagaagcggtcacaagcaattattctgttcccaatg 460
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 557 ttctcatgtctccaatttagagcgctgtggaatagccacattccagactccagctg 616
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 461 tctcaagacgaataagctgttcgggtgtttatgtcttccctgtccagaggtgcaagctg 520
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 617 attaaattgcttgcgcagcttctggaagcttcaacccagaggaatagttacacact 676
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 521 acctcaactgcatggaagactgtgtccagagctcagagtcgagtggaattacttccgaata 580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 677 tgtgtggcagaagatttccctcgaagtcacaaatttgatgtgtcagagttaaaaaac 736
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 581 catgtggagcagacttccctataaagagcaatgacagatggtccagagctctaagaatgt 640
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 737 tcaatggagacaatatgttggagacggttgaaacccccaacagtaaatggaagattca 766
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 641 tgaatggaggaatagatgagtcagagtgactcctctaagcacaagaacccgcgtgga 700
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 797 cttaacatcagaaacttaagacgggtgccttaagaaatgtgaagctacacaataaggaca 856
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 701 aatatcaacttgagtgagtgagagacacattaac-----ctaacca 742
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 857 acatctccaaggaagacaccccccacataacattcagatattgttggcagtgctatttg 916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 743 acaagaagaagagatcccccctataatataactatgattacaggaatgctgacattg 802
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 917 tttaagtcagaactgttgaatataatttcaacaactccatcgttcaagactttttg 976
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 803 tggcttcccgagatttgcctcaacatgttttgagaagccctaataccacaacactggt 862
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 977 cctgtcttaagaagacatactcctctatgaagacatttggtaacttattgctggttc 1036
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 863 aatgggcaaaagacactataagcccaagatgaacccctctggtgcccacttccagctgac 922
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1037 cagaataactcgtggagagattccagatcagcccaagatg---tgctgcatctgcagatga 1093
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 923 ggttgatgctcgtctgttcccaacacacccaagtagcagacatcagatgacttcta 982
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1094 agactcgcctgtcaagtgtgaattactatgaaggcttcttctatccag----- 1143
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 983 ttgcagagctgtgtcaagtgtcagaggtcatgaggaacatcgataaagggtgctcttattg 1042
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1144 -----tgactggaatcactcagacgtgtgtattatgtgagctgcagaaattaaagt 1198
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1043 ctccctgctcgtgaatccaccagcggtctatctgtttatgggagctgggactgaatt 1102
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1199 ggcctacaaagatggagactggttggtaataatttgatttcaagtgtaccctatct 1258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  
```

Db 1103 ggaatgcttaaaacaccatccactgttgcacacaagttgaccacaaggtagatgataatg 1162
 QY 1259 tgatnaatgcttgcagagaagct 1283
 Db 1163 cctctcagtgcttaagaagaataact 1187

```

RESULT 4
AA96569
ID AA96569 strand: DNA; 1317 BP.
XX
XX AAA96569;
XX
XX 08-FEB-2001 (first entry)
XX
XX DNA encoding a core 2 beta-1,6-N-acetylglucosaminyltransferase.
XX
XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
XX cancer; cardiovascular disorder; inflammatory disorder; asthma;
XX rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
XX septic shock; adult respiratory distress syndrome; ARDS; cancer;
XX platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
XX clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
XX diverticulitis; ulcerative colitis; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1317
XX FT /*tag= a
XX FT /*product= "core 2
XX PN CA2296936-A1.
XX
XX 03-AUG-2000.
XX
XX 03-FEB-2000; 2000CA-2296936.
XX
XX 03-FEB-1999; 990US-0118674.
XX
XX (GLYC-) GLYCODESIGN INC.
XX
XX Korczak B, Lew A;
XX WPI; 2000-59476/57.
XX P-PSDB; AAB18995.
XX
XX New nucleic acid molecules of core 2
XX beta-1,6-N-acetylglucosaminyltransferase useful for providing new
XX compositions for treatment of disorders mediated by the enzyme
XX including cancer, cardiovascular and inflammatory disorders.
XX
XX Claim 4: Page 50; 66pp; English.
XX
XX The present sequence encodes a human core 2
XX beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
XX The polypeptide can be used to treat diseases and disorders, such as
XX cancer, cardiovascular disorders and inflammatory disorders including
XX asthma, rheumatoid arthritis, inflammatory bowel disease,
XX arteriosclerosis, septic shock, adult respiratory distress syndrome
XX (ARDS) and cancer. Various platelet-mediated pathologies such as
XX atherosclerosis and clotting can also be treated. The polypeptides of
XX the invention are predominantly expressed in gastrointestinal tissue
XX (stomach, colon, intestine, testis) and are elevated in cancer.
XX Gastrointestinal disorders that may be prevented or treated include
XX ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
XX and ulcerative colitis. The antibodies may be used in
XX immuno-histochemical analysis, to detect the novel polypeptide and to
XX localize it to particular cells and tissues and to specific subcellular
XX locations and to quantitate the level of expression.
XX
XX Sequence 1317 BP; 359 A; 307 C; 322 G; 329 T; 0 other;
SQ
  
```

Query Match 14.1% Score 191.8; DB 21; Length 1317;
 Best Local Similarity 52.5%; Pred. NO. 5.9e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

QY 317 tgaccagtgatgtgacattatcaagactcaagaggtlaagctcaaaagctgtctcaa 376
DB 317 tcaccagagactgtgagacttcaagctgaagaaagttcatacagttcccaagagca 376
QY 377 agggaggaanaagcttcccaatagccttcttgggtgtcccaagaagtcaatttgg 436
DB 377 aagaagaggtgaggttccctatgtatctcattatggttattcagagaagaattgaact 436
QY 437 ttgaagagcttatccatgtatatacaaccagacaatatttactgtgcatcattatgc 496
DB 437 ttgaagagctatcgtgagctgtgtatgcccctcagaacatactactgtgtcattgtgatg 496
QY 497 gtaagagcaccgtatcccttcaaaagttgcacagaaacttgcgaagtgtctcccaata 556
DB 497 agaagtcaccagaaacttcaaaagagcggtcaagaacattattcttgcctcccaatg 556
QY 557 ttctatgtctcccaatlaagagcggtggaataatgcccacattccagactccaagcg 616
DB 557 tctctatagccagtaagctgtgtcggtgtttagtgcctctgtccaggggtgcagaagctg 616
QY 617 atttaattgtctgtcggaactcttgaagcttcaatccagttggaanaattgtatcaact 676
DB 617 acctaacctgcataggaagactgtctccagagctcagtgccgtggaaacttccctgaata 676
QY 677 tgtgtggcagaagatttccctcgaagtcacaaatttgaattgtgtcagaagttgaaaaaac 736
DB 677 catgtggagcagacttctcctataaagagcattgcagagatgttccaggtcttcaagagt 736
QY 737 tcaatggagcaaatctgtgtgagacgtgtgaaccccccaacaagttaattgaaagattca 796
DB 737 tgaatggagagatagcatgtgagtcagagatccctcacaagaacaaacccgcttga 796
QY 797 cttaaccatcaacttaactagacgggtgctttagaatgtgaagctacataaagaacaa 856
DB 797 aatatcactttaggttagtagagacacattacac-----ctaaaca 838
QY 857 acatctccaaggaagacccccccataacatcagatatgtgtggcagttcctattttg 916
DB 839 acaagaagaagatccctcccttataatttaactatgttcaaggaattcgtacatgt 898
QY 917 tttaagcaaacatttcttaatatatttcaacaactcatgcttcaagaactttttg 976
DB 899 tggctcccgagatttgcaccacatgtttgaaagacccttaaccccaacaacatgattg 958
QY 977 cctgtctaaagacacatactctcctgtatgacacttttgggtacttgaattcgggttc 1036
DB 959 aatgggtlaaagacactatagccagatgaacacctctgggcccacttccagcgtagac 1018
QY 1037 caggaatcctgtggagatttccagatacagcccaagatg---tgtctatctgcagata 1093
DB 1019 ggtgtgagctgtgctgttcccaaccacccaagaaagtaagacatcccaagatctta 1078
QY 1094 agacacgccttctcaagtggaattactatgaaggttttcttccatccaggt----- 1143
DB 1079 ttgcagagctgtgtcaagtgtgacgggtcatgtagggagacatcgataaggtgtccttatg 1138
QY 1144 -----tgtactgtatcactccttgaaagcggtgttatttaaggactccaagaattaaagt 1198
DB 1139 ctccctgctctggaatccaacagcggtcatctgttlaagggctgtgggaactgaatt 1198
QY 1199 gcttatcaaaagatgacactgtgttctgaataaatttgatctcaagttgacaccttact 1258
DB 1199 ggaatgcttaaaacacatacaccctgttgccaacaagtttgacccaagaagtgatagttaagt 1258
QY 1255 tgatttaaaagcttgcagaanaagct 1283
DB 1259 ctcttcagtgcttgaagaataact 1283

```

RESULT 5
 AAA96570
 ID AAA96570 standard; DNA; 2108 BP.

```

XX AC AAA96570;
XX AC AAA96570;
XX DT 08-FEB-2001 (first entry)
XX XX
DE A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.
XX XX
XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
XX cancer; cardiovascular disorder; inflammatory disorder; asthma;
XX rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
XX septic shock; adult respiratory distress syndrome; ARDS; cancer;
XX platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
XX clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
XX diverticulitis; ulcerative colitis; ss.
XX OS Homo sapiens.
XX XX
XX CA2296936-A1.
XX XX
XX 03-AUG-2000.
XX PD
XX PF 03-FEB-2000; 2000CA-2296936.
XX PR
XX PR 03-FEB-1999; 9905-0118674.
XX XX
XX PA (GLYC-) GLYCDESIGN INC.
XX PI
XX PI Koczak B, Lew A;
XX DR
XX DR MPI; 2000-594746/57.
XX XX
XX New nucleic acid molecules of core 2
XX PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
XX PT compositions for treatment of disorders mediated by the enzyme
XX PT including cancer, cardiovascular and inflammatory disorders.
XX PT
XX PS Claim 4; Page 51-52; 66pp; English.
XX PS
XX XX
XX The present sequence encodes a partial human core 2
XX CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
XX CC The polypeptide can be used to treat diseases and disorders, such as
XX CC cancer, cardiovascular disorders and inflammatory disorders including
XX CC asthma, rheumatoid arthritis, inflammatory bowel disease,
XX CC arteriosclerosis, septic shock, adult respiratory distress syndrome
XX CC (ARDS) and cancer. Various platelet-mediated pathologies such as
XX CC atherosclerosis and clotting can also be treated. The polypeptides of
XX CC the invention are predominantly expressed in gastrointestinal tissue
XX CC (stomach, colon, intestine, testis) and are elevated in cancer.
XX CC Gastrointestinal disorders that may be prevented or treated include
XX CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
XX CC and ulcerative colitis. The antibodies may be used in
XX CC immuno-histochemical analysis, to detect the novel polypeptide and to
XX CC localize it to particular cells and tissues and to specific subcellular
XX CC locations and to quantitate the level of expression.
XX XX
XX Sequence 2108 BP; 596 A; 464 C; 509 G; 539 T; 0 other;

```

Query Match 14.1% Score 191.8; DB 21; Length 2108;
 Best Local Similarity 52.5%; Pred. NO. 7.2e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

QY 317 tgaccagtgatgtgacattatcaagactcaagaggtttagtccaaaagctgtctcaa 376
DB 584 tcaccagagactgtgagacttcaagctgaagaaagttcatacagttcccaagagca 643
QY 377 agggaggaanaagcttcccaatagccttcttgggtgtcccaagaagtcaatttgg 436

```

```

Db 644 aagaagaggtgaggtccctatgcatcactatggtgatcattgaagaagtgtgaaact 703
QY 437 ttgaagagcttatccatgcttatatacaacccagacataatttaccgcatcattatgctc 496
Db 704 ttgaagagcttatccatgcttatatacaacccagacataatttaccgcatcattatgctc 763
QY 497 gtaaggacactatccatccttcaaaagtgtccatgaacaatttgaagtgtcttcccaata 556
Db 764 agaagtcctccagaaactttcaaaaggcggtcaaaagcaatttattcttcttcccaaatg 823
QY 557 ttcttaagcttcccaataatgagctgtgtggaataagccacatttccgaactccaagctg 616
Db 824 ttcttaagcttcccaataatgagctgtgtggaataagccacatttccgaactccaagctg 883
QY 617 atttaattgtcttgcgaacttgcgaacttgcgaacttgcgaacttgcgaacttgcgaact 676
Db 884 acctcaactgcataagagacttgcctccagagctcagtgccgtggaataattcttccgaata 943
QY 677 tgtgtggagcaagatttccctcgaagtcgaatttgaattgtgtcagagttgaaataac 736
Db 944 catgtggagcaagatttccctcgaagtcgaatttgaattgtgtcagagttgaaataac 1003
QY 737 tcaattggagcaaatatgttggagacggtggaaccccccaacagtaaatgtgaaagattca 796
Db 1004 tgaattggagcaagatttccctcgaagtcgaatttgaattgtgtcagagttgaaataac 1063
QY 797 cttaacatcatgaacttgaagtcgaagtcgaatttgaattgtgtcagagttgaaataac 856
Db 1064 aattacacttctgagtgagtgagagacattcac-----ctaacca 1105
QY 857 aacttcccaaggaacaccccccaatacatcagataattgttggcagtgcttatttg 916
Db 1106 acaagaagagagatccctcccttataatttactatgttttaccagtgagtgatcatg 1165
QY 917 tttaagtcagacatttcttaataatattttcaacaactccatgcttcaagaacttttg 976
Db 1166 tggcttccgaagatttgcctcacacatgttttgaagaaccccttaacccaacactgattg 1225
QY 977 cctgtgtaaaagacacatactctctcgtatgagacatttggcagcttgcagtgctc 1036
Db 1226 aatgggttaaaagacacatactctctcgtatgagacatttggcagcttgcagtgctc 1285
QY 1037 cagggaactctgtggagatttccacatccagcccaagatg---tgcctatctgcagata 1093
Db 1286 ggtgtgagctgtctctgttcccaaccaccccaaglaagacatcccaacatgacttcta 1345
QY 1094 agactgccttgcagtggaattactatgaagcttttctatccagctt----- 1143
Db 1346 ttgcccagctgtgtaagtgagcaggttcagagagagagacatgataagggtgctcctatg 1405
QY 1144 -----tgcctgtgactcctcctcgaagcggtgtatattatggaactcagaaattaaagt 1198
Db 1406 cctcctgtcctgtgaatcccccagcgaggtatctgcgttattatggtgtggagcttgaatt 1465
QY 1199 ggcattatcaaatgtgacatgttctgtcctaataatgtatcattcaaggtgagccatct 1258
Db 1466 ggaatgttcaaaacacatccactgttgccacacagatttgcacccaaggttagtgaatg 1525
QY 1259 tgattcaaatgtcttgcaagaagct 1283
Db 1526 cctctcagtcgtcttgaagaataact 1550

```

```

RESULT 6
AAC62134
ID AAC62134 standard: DNA: 2128 BP.
XX AAC62134;
XX
XX 06-MAR-2001 (first entry)
XX
XX Nucleotide sequence of beta-1-6-N-acetylglucosaminyltransferase.
DE
XX

```

```

KW Human: beta-1-6-N-acetylglucosaminyltransferase; C2GNT-M; inflammation;
KW membrane protein; branched sialyl Lex; L-selectin; immune reaction;
KW inflammation; tissue rejection; tumour metastasis; SS.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 354..1670
FT /tag= a
FT /product= "beta-1-6-N-acetylglucosaminyltransferase"
FT polyA.signal 2100..2105
FT /tag= b
XX
XX US6136580-A.
XX
XX 24-OCT-2000.
XX
XX 19-JAN-1999; 99US-0233506.
XX
XX 19-JAN-1999; 99US-0233506.
XX
XX (BURN-) BURNHAM INST.
XX
XX Fukuda M, Yeh J;
XX
XX WPI: 2001-040238/05.
XX
XX P-PSDB; AAB30518.
XX
XX New C2GNT-M polypeptides having core 2, core 4 and I branching
XX PT beta-1-6-N-acetylglucosaminyltransferase activities for preparing
XX PT reagents useful for diagnosing, preventing or treating inflammation or
XX PT tumour metastasis
XX
XX PS Example 1; Fig 4; 25pp; English.
XX
XX CC The present sequence encodes a human
XX CC beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4
XX CC and I branching activities. It is designated C2GNT-M. C2GNT-M is a
XX CC membrane protein that is predominantly expressed in colon, small
XX CC intestine, trachea, stomach and thyroid, as well as in certain cancer
XX CC cell lines. C2GNT-M polypeptides may be used to prepare molecules having
XX CC highly branched sialyl Lex and L-selectins, which may be subsequently
XX CC used to modulate immune reactions, e.g. inflammation and tissue
XX CC rejection, and to prevent or inhibit tumour metastasis.
XX
XX SQ Sequence 2128 BP; 569 A; 477 C; 526 G; 556 T; 0 other;

```

```

Query Match 14.1%; Score 191.8; DR 22; Length 2128;
Best Local Similarity 52.5%; Pred. NO. 7.3e-44;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

```

QY 317 tgacagtgattgtgacattatcaagacttaagagttatgtctcaaaagctgtctcaa 376
Db 670 tcacacagagctgtgagcacttcaagctgaagaggttcataacagttccactgagca 729
QY 377 agggagagaaagtgcttccatagcctatcttgtgtgtgcccaagaagatgacatg 436
Db 730 aagaagaggtgtgagcttccatgtgacatccctatgtgattcattgaaatgaaact 789
QY 437 ttgaagagcttatccatgcttatatacaacccagacataatttaccgcatcattatgctc 496
Db 790 ttgaagagcttatccatgcttatatacaacccagacataatttaccgcatcattatgctc 849
QY 497 gtaaggacacttgaacttcaaaagtgtccatgaacaatttgaagtgtcttcccaata 556
Db 850 agaagtcctccagaaactttcaaaaggcggtcaaaagcaatttattcttcttcccaaatg 909
QY 557 ttcttaagcttcccaataatgagctgtgtggaataagccacatttccgaactccaagctg 616
Db 910 ttcttaagcttcccaataatgagctgtgtggaataagccacatttccgaactccaagctg 969
QY 617 atttaattgtcttgcgaacttgcgaacttgcgaacttgcgaacttgcgaacttgcgaact 676

```

Db 970 acccaactgcatggaagatgtgtccagagctcagagtcggtggaatacttctgaata 1029
 QY 677 tgtgtggcaagatcttccctgaagtcgaatttgaattgtgtcgaatttgaataac 736
 Db 1030 catgtggaaggaatttcccttaagaagcaatgacagatggttccagctctaaagtgt 1089
 QY 737 tcaatggagcaaatgtgtgagacggtgaaaccccccaacaaatgaattgaaagattca 796
 Db 1090 tgaatggaggaataagatagatgagtcagagtaactctcctaagacacaaagaaacccgctgga 1149
 QY 797 ctaccatcatgaacttaagacaggtgccttatgaattgtgaagctacataagagcaa 856
 Db 1150 aatccaacttgaggtagtgagagacacattacac-----ctaacca 1191
 QY 857 acatctcaaggaagaccccccacataacattcgaatttgttggcagtgctattttg 916
 Db 1192 acaagaagaagagatcccccctataatttaactatgtttacaggaatgctgaattg 1251
 QY 917 tttaagtcgaagcatgtgttaataatatttcaacaactccatcgttcaagactttttg 976
 Db 1252 tgcctcccgagattcgtccaacatgtttgaaagacccctaaatcccaactgattg 1311
 QY 977 cctgtctaaagacacatactctctctatgagacatttgggtcacttgatcgggttc 1036
 Db 1312 aatgggtaaaagacactatagcccaatgaacacccctgggacacccctcagcgtgcac 1371
 QY 1037 caggaaactcgtgggaatttccagacacagccagagtg---tgctcgtactgcaagata 1093
 Db 1372 ggtgtgactcgtcgtctgtctccacacaccccaagtaagcactctcagatgaattcta 1431
 QY 1094 agactcgcctgtcaagtgtgaattacatgaagcttttctacaccagt----- 1143
 Db 1432 ttgcagagcgtgtcaagtgagaggtatgaggaagacatcgaaaggtgtccttaag 1491
 QY 1144 -----tgatcgtgatacctcgaagcgtgtatattatgtgagcgcagcaatgaagt 1198
 Db 1492 ctccctgctcgtgatacaccacagcggtatcgtgttatggcgctggggaactgattc 1551
 QY 1199 ggcattacaagaatgagacattgtgtctataaattgattcttaaggtggaacctatc 1258
 Db 1552 ggaatgctccaacacatcacctgtgtgccaagaattgaccccaaggtatgataatg 1611
 QY 1259 tgattaatgcttgacagaagaact 1283
 Db 1612 ctcttcagtgcttagaagaataact 1636
 RESULT 7
 AAC99109 standard; cDNA: 2229 BP.
 ID AAC99109:
 AC AAC99109:
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.
 XX
 KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200055320-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05989.

XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 DR WPI: 2000-579444/54.
 XX P-PDSB: AAB54344.
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 XX
 PS Claim 1; Page 759-760; 1379pp; English.
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;
 Query Match 14.1%; Score 191.8; DB 21; Length 2229;
 Best Local Similarity 52.5%; Pred. No. 7.4e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;
 QY 317 tgaccagatgattgacattatcagacrtcgaaggtatgtctcaaaagctgtgtcaca 376
 Db 743 tcaccagagactgtgagcacttcaagcgtgaaggtacatcacgttccactgagca 802
 QY 377 aggaagagaaaagcttcccaatagcctattcttgggtgtccaaaagatgcaattatg 436
 Db 803 aagaagaagtgaggttccctatgtcactatcgtatgtgtatcagagaatgaaact 862
 QY 437 ttgaaggtctatccatgtatatacaaccagacataattacgtacatcatatgac 496
 Db 863 ttgaaggtctacgcgaggtgtgtatgcccctcaagaacataactgtgtcacaatgtgagtg 922
 QY 497 gtaagcactgtactcctcaaaagtgtccatgaacaattagtaagtgtcttccaca 556
 Db 923 agaaagcccccagaacttccaagaagcgtgtcaagaacattatcttcttccccaag 982
 QY 557 ttctcatgtctccaattagaggtgtggaatgtccacatttccagaactcagagctg 616
 Db 983 tctctatagcagatagcgtgttcggtgtgtatgtatccctccgtgtccaaggtgacgctg 1042
 QY 617 atttaattgctgttcgacacttgaagcttccatccacatcagtggaatgttatcaact 676
 Db 1043 accccaactgcatggaagacttgcacagactccaagtgcgtggaatcttccogaata 1102
 QY 677 tgtgtggcaagaatttccctgaagtcgaatttgaattgtgtcagaattgaaataac 736
 Db 1103 catgtggaaggaatttccctataagaagcaatgacagatggtgcaggtctcacaagatgt 1162

```

QY 737 tcaatgagacaataatgttggagagcgtgtaaaaccccaacagtaattggaagatcca 796
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1163 tgaatggagagaataatgataatgagatcagagtgtaacctccaaacaaaccccgctgga 1222
QY 797 ctaccacataatgaacttgaacgggtgcctttgaatatgtgaagctaccataaagacaa 856
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1223 aatatacatttggagtagtagagacaataaac-----ctaacca 1264
QY 857 acatctccaagaagacaccccccaataacattcagatatgttggcgagtgctattttg 916
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1265 acaagaagaagagatcctcccttataatttaactatgttaacaggaatgctgacattg 1324
QY 917 tttaagccaagcatttcttaataatatttcaacaacatccatcggtccaagctttttg 976
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1325 tggctcccgagatttgcacacatgttttgaagaaccccaaaccccaaacctgattg 1384
QY 977 cctggctcaagaacacatactcctctgtagagcattttggggtctgacttgatcggttc 1036
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1385 aatgggttaagaagacattatagccagatgtaacacactctggccaccccttaagcgtgac 1444
QY 1037 caggaatacctggggagagatttccagataagccagagatg---tgctgactgacagta 1093
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1445 gttgagatcgtcgtctctgtcccaaccccaacagtaacagatctcagacatgacttcta 1504
QY 1094 agactcgccttgcagaaggaattacttgaaggcttttctatccagct----- 1143
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1505 ttgccaagctcgtcagaagcaggggtcatgagggagacatcgataaagggtctcctcattg 1564
QY 1144 ----tgtaactgatactcacttcgaagcgtgtatttattgaagctgcagaattaaagt 1198
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1565 ctccctgctcgtgaatccacacagcggtcatctgtttatgggctggggagatttgaatt 1624
QY 1199 ggcctataaagatggaattggttgcataataatttgatttcaaggtggaacctatct 1258
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1625 ggaatctccaacaacatcaactcgttgcacaagaatttgacccaagaagtagatataatg 1684
QY 1259 tgattaaatgcttgcagaaagaact 1283
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1685 ctcttcagtgcttagaagaataacct 1709
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 8

AAH34463 ID AAH34463 standard; cDNA: 2236 BP.

AAH34463; AC

03-SEP-2001 (first entry) DT

Human colon cancer antigen encoding cDNA SEQ ID NO:1545. DE

Human; colon cancer; colon cancer antigen; diagnosis; detection; KW

colorectal carcinoma; chromosome 15; ss. XX

Homo sapiens. OS

WO200122920-A2. PN

05-APR-2001. PD

28-SEP-2000; 2000WO-US26524. PF

29-SEP-1999; 99US-0157137. PR

03-NOV-1999; 99US-0163280. PA

(HOMA-) HUMAN GENOME SCI INC. PI

Ruben SM, Barash SC, Birse CE, Rosen CA. DR

WPI: 2001-235357/24. DR

P-PSDB; AAG75058. XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides.
 useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 1; Page 3190-3191; 9803pp; English.
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 2236 BP; 609 A; 503 C; 548 G; 574 T; 2 other;

Query Match 14.1%; Score 191.8; DB 22; Length 2236;
 Best Local Similarity 52.5%; Pred. No. 7.4e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

QY 317 tgaacagtgatttgaacatttaccagctcctaaagaggtatgctcaaaagcttgcctaa 376
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 750 tcccaagagactgtgagcacttcaagcttgaagaaggtacatccactgagca 809
QY 377 aggaagagaagaagcttcccaatagcctatccttggtttcccaagaagtgcaattatg 436
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 810 aagaagagtggtgaggttcccatgcatctctctgttgaatttcaaggaagtgaataact 869
QY 437 ttgaagagcttaccatgctatatacaacacagacaatttaccgacattatgac 496
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 870 ttgaagagctactgcagcagctgtgtatgcctcccaacacatacttgcctatgtagtg 929
QY 497 gtaagcactgtatccttcaagaagttgcataagcaatttgaactgaagcttcccaata 556
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 930 agaagtcocccagaacttccaagaagcgggtccaagaacattatcttcttcccaaatg 989
QY 557 tttaactgcttccaatagagcgtgtgaatatgcccacatttccagaactcagctg 616
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 990 tcttcataagccaatgaactgtgttcgggtgtttatbctccctcgttccaagggtccaagt 1049
QY 617 atttaaatgcttgcgaacttgaagttcctcaatcccaagtggaataatgttataact 676
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1050 acctcaactgcataaggaacttgcctcagaagctcgaatgcgttggaaatcttccgaata 1109
QY 677 tgtgtgggcaagaatttccctgaagccaatttgaatttgaattgtcagaagtgtgaaaaac 736
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1110 catgttggaagcacttccataaagaagcatcagaagatgttcagagcttccaagatgt 1169
QY 737 tcaatgagacaataatgttggagagcgtgtaaaaccccaacagtaattggaagatcca 796
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1170 tgaatggagagaataatgataatgagatcagagtgtaacctccaaacaaaccccgctgga 1229
QY 797 ctaccacataatgaacttgaacgggtgcctttgaatatgtgaagctaccataaagacaa 856
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1230 aatatacatttggagtagtagagacacattacac-----ctaacca 1271
QY 857 acatctccaagaagacaccccccaataacattcagatatgttggcgagtgctattttg 916
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1272 acaagaagaagatcctcccttataatttaactatgtttcaagggaatgctgacattg 1331
QY 917 tttaagtaagaagatttgaataatataatttcaacaaccccaactcgttccaagcattttg 976
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 1332 tggctccgcgagatctgcgtccacacatgcttttgagaccccttaactcccaacatcgtatg 1391
 QY 977 cctggtcctaagacacatcactctcccgatgagcactttcggtcactctgattcggttc 1036
 Db 1392 aatggttaaaagacactatagcccgatgaacactctgggaccccttcagcggtcac 1451
 QY 1037 caggaatcctgggagattccagatccagccagatg---gtctgcttcagagta 1093
 Db 1452 ggtgagatgctgctcgtctcccaacaccccaagtaagacactcagacatgacttcta 1511
 QY 1094 agactcgcttgcgaagtgaattactatgaaggttttctatcccaat----- 1143
 Db 1512 ttgcagagctggtcaggtgaggtatagggagacacgtataggtggtccttatg 1571
 QY 1144 ----tctactggtcctcacccttcgaagcggtgtgtatttatggagctgcgaattaggt 1198
 Db 1572 ctccgtctctgtaatccaccacagcggtctctcgtcttctatgggctggggaacttaatt 1631
 QY 1199 ggcattacaagaatggaactgtgttgcataataattgaattcctaaggtggacctatct 1258
 Db 1632 ggaatgcttcaaacaccatcacctgttgccacaagttgacccaaggtatgataatg 1691
 QY 1259 tgattaatgcttgcgcgcaaaagct 1283
 Db 1692 ctctcagtgcttagaagaatacct 1716

RESULT 9
 ID AAA48623
 AC AAA48623:
 DT 19-SEP-2000 (first entry)
 DE Human C2/4Gnt CDNA.
 XX Human: C2/4Gnt; UDP-N-acetylglucosamine: O-glycan biosynthesis;
 KM O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 496..1812
 FT /*tag= a
 FT /product= "C2/4Gnt"
 FT primer_bind 586..605
 FT /*tag= b
 FT primer_bind 1794..1813
 FT /*tag= c
 FT polyA_signal 2244..2249
 FT /*tag= d
 XX
 PN WO200034449-A2.
 XX
 PD 15-JUN-2000.
 XX
 PF 03-DEC-1999; 99WO-DK00677.
 XX
 PR 04-DEC-1998; 98DK-0001605.
 XX
 PA (CLAU/) CLAUSEN H.
 XX
 PI Clausen H, Schwientek T;
 DR WPI; 2000-423407/36.
 DR P-PSDB; AAY94492.
 DR
 XX New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
 PT probe for the detection of specified glucosaminyltransferase from
 PT other species and related organisms
 XX

PS Claim 5; Fig 2; 47pp; English.
 CC The present sequence encodes human UDP-N-acetylglucosamine:
 CC N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
 CC (C2/4Gnt). The protein is the third member of the family of O-glycan
 CC beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
 CC based O-glycans on to oligosaccharides, glycoproteins and
 CC glycosphingolipids. C2/4Gnt can therefore be used in the production of
 CC appropriately glycosylated glycoconjugates with particular enzymatic,
 CC immunogenic, or other biological or physical properties. The nucleotide
 CC sequence is useful as a probe for the detection of C2/4Gnt from other
 CC species and related organisms and for the recombinant production of
 CC C2/4Gnt polypeptide. The nucleotide sequence was identified by analysis
 CC of EST database sequence information. Oligonucleotides derived from EST
 CC clone 17865 of ATCC were used to isolate two full-length C2/4Gnt clones
 CC from a human foreskin genomic pl library by 5' RACE PCR. RT-PCR was
 CC performed using Colo205 human cell line mRNA in order to produce cDNA
 CC for expression of C2/4Gnt in Sf9. The control of O-glycan core assembly
 CC has been implicated in tumour progression and metastasis.
 XX
 SQ Sequence 2319 BP; 615 A; 528 C; 559 G; 617 T; 0 other:

Query Match 14.1%; Score 191.8; DB 21; Length 2319;
 Best Local Similarity 52.5%; Pred. No. 7.5e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagtgatgtgacattatcagactctaagaggttatgtcctaagaagctgtgtccta 376
 Db 812 tcaccagagactgtgagcacttcaagctgaggaaggttcatagatgccactgagca 871
 QY 377 aggaagagaagaagcttcccaatagcctattcttgggtgtccacaagaatgcaattatgg 436
 Db 872 aagaagaggtgagatccctatgcatcactatcgtatgcatcagagaagatttaaacct 931
 QY 437 ttgaaggttcatcagatcatatatacaaccagacacatatttactgtcatcattatgctc 496
 Db 932 ttgaaggttactgagctgtgtgtgtatgccccagacaatactatgctcatgtgtgctg 991
 QY 497 gtaaggaactgtatcctcctaaggttgcatagaataatttgctaatggtcttcccaata 556
 Db 992 aagaatccccagaacttccaagaagcggtcaagaacattattctgtcttcccaatg 1051
 QY 557 ttccattgtcttccaatataagagctgtggaatagcccaattccagatccagctgctg 616
 Db 1052 tcttcattagccagatagctgtgtgtgtgttattatgcttccctgtgcccaggttgcagctg 1111
 QY 617 atttaattgctgtcggacacttctgaagcttcaatcccaagtgtgaatatgttatcaact 676
 Db 1112 acctcaactgcatggaagactgtctccagagctcagtgccgttgaataacttccatgaata 1171
 QY 677 tgtgtggcaagatttcccccgaagtcgaatttgaattgtgtgcagagttgaaataaac 736
 Db 1172 catgtgggagcgaacttccctataaagagaacagcagagatgttccagctctcaagaagt 1221
 QY 737 tcaatggagcaaatatgttggagcaggtgaaaccccaacaacagtaaatgtgaaagattca 796
 Db 1232 tgaatgggagagatagcagtgaggttcaagagttcccttaagcacacaaccccgctgga 1291
 QY 797 cttaacatcatgacttagcaggtggtccttatgaaatgtgagtgagtcacaataagagcaa 856
 Db 1292 aatatcaactttaggtatgtgagagacattcac-----ctaaacca 1333
 QY 857 acatctcaagaagaagaccccccataacatcagatatgttggcagtgcttatttg 916
 Db 1334 acaagaagaagagatcctcccttataatttaactatgtttacagaggaatcgctacattg 1393
 QY 917 tttaagtaacgattgtttaaataattatttcaaacacccatcgcttaagactttttg 976
 Db 1394 tggcttccggaatttgcaccaactgttttgaagaacctaataccacaactgattg 1453
 QY 977 cctgtctaaagacacatactccttgatgagcaatttgggtacacttgattcgggttc 1036


```

RESULT 11
AA116201
ID AA116201 standard; cDNA; 1807 BP.
XX
AC AA116201;
XX
DI 22-APR-1996 (first entry)
XX
DE Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme cDNA.
XX
KW Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme; IGT;
KW glycosyltransferase; blood group; I antigen; polylactosaminoglycan;
XX ds.
XX Homo sapiens.
XX OS
XX FH Key Location/Qualifiers
XX FT CDS 255..1457
XX FT /*tag= a
XX
PN US5484590-A.
XX
PD 16-JAN-1996.
XX
PF 09-SEP-1993; 930S-0118906.
XX
PR 09-SEP-1993; 930S-0118906.
XX
PA (LJOL-) LA JOLLA CANCER RES FOUND.
XX
PI Bierhuizen MFA, Fukuda M;
XX
DR WPI: 1996-087019/09.
XX
DR P-PSDB: AAR92474.
XX
XX
PT Human beta-1,6-N-acetylglucosaminyl transferase, I-branching enzyme
PT - used to develop prods. for the study, detection and treatment of
PT pathological conditions involving the enzyme.
XX
PS Example 4; Column 29-32; 29pp; English.
XX
CC A cDNA insert (AA116201) in pCDNA1-IGT codes for human beta-1,6-N-
CC acetylglucosaminyltransferase, I-branching enzyme (IGT) (AAR92474),
CC an enzyme that converts blood group I antigen to I antigen. The
CC cDNA was isolated from a cDNA library pred. from human PA-1
CC teratocarcinoma cells in vector pCDNA1 following transfection into
CC CHO-Py-leu cells and screening with human anti-I antibodies. The
CC cDNA can be used to regulate the expression of human IGT or to
CC modify its biological function, to produce soluble or membrane-bound
CC forms of IGT in host cells, to breed transgenic animals, and to
CC design antisense oligonucleotides.
XX
SQ Sequence 1807 BP; 511 A; 401 C; 398 G; 497 T; 0 other;

```

```

Query Match 13.6%; Score 185.2; DB 17; Length 1807;
Best Local Similarity 52.1%; Pred. No. 4,9e-42;
Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```

```

QY 327 ttgtgacattatcagacttaagaggtatgctcaaaagctgtctcaaaagagagaa 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 467 ttgcaagaaatacttgaccacagcaccatcacacagcccttatactaaagaagc 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 aagcttccaatagcctattcttggtgtccaaagtgatgataatgttggaagct 446
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 527 tgccttcccttgacataataatggtcaccctcctccttgacaccttgcaagcct 586
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 447 tatccatgtatataaacacgaacataatttcatcattatgatcgtgaagcacc 506
   || |||| |||| || |||| |||| || |||| |||| || |||| |||| ||
DB 587 cttcagagctattacatgccccaataatactctgttcatcattgtatgaaagaagc 646
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 507 tgatacctcaaaagttgcacatgaacaattagcttaagtctctccaataatttctatgc 566

```

```

DB 647 aactgaattlaaagatcggtagagcaacttaagctgtctcccaaacgcttcttcgac 706
   || |||| ||| ||| ||| ||| |||| ||| ||| ||| ||| ||| |||
QY 567 ttccaattagaagctgtggaatatgcccacattccagactccaggtcatttaattg 626
   |||| ||| ||| |||| |||| |||| |||| |||| |||| |||| |||| ||
DB 707 ttccaagatgaaccgctgtctatgagggatctccaggtccaggtcgaacctg 766
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 627 ctgtcgacctctgaagcttcaatccagtggaataatgttatacaactgtgtggca 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 767 catcagaatcttctgctcagagctcctagaagtaactgtatcaacaccttgaggca 826
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 667 agatttcccttgagaagtaaatlttgatgtgtgtcagagttgaaaaaactccaatggagc 746
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 827 agacttccccctgaaaacacaagaagaaatagtcacatctgaagaagatttaagtaa 886
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 747 aaatatgttggagacgttgaaacccccaacagtaaatgtggaagattcattaccatca 806
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 887 aaatacaccacaggggtgctgcgccccagctcactgaattgagcgaactaataatgtcca 946
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 807 tgaactiagaggggtgctctatgatatgtgaagctacccaataagggacaacatctccaa 866
   || || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 947 ccaagagacacttgagca-----aagagcttccatgtatgataagaacaacagcgttgaa 1000
   || || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 867 ggaagcaccgcccaataacattcagataattgttgcaagtctatttgtttaagta 926
   || |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1001 accgctcccccataatctcacaatttacttggctgtgctatgtgtgtctatcaag 1060
   || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 927 agcatttgttaatatatttcaacaactccatcgttccaagacttttgcctggtctaa 986
   || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 1061 agagtttgcacacttgttctgcataagccacgggcgttgattgtctcagtggtccaa 1120
   || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 987 agaacatcactctcctgatgagcacttlttggtcacttgatccgggttcacagaatacc 1046
   |||| ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
DB 1121 ggacacttcaagtcctgatgagcattctcgtgtgacactcaatagatgattccaggtgtcc 1180
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
QY 1047 tgggagagttccacagacagccacagagatgtgtcgtacatcgagagtaagactgcctgt 1106
   || || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1181 ttgctctgtccaaatgatgtctcggactg-----gaaacctcagaagctat 1225
   || || || || || || || || || || || || || || || || || || || ||
QY 1107 caagtgaattactacatgaagcttctctatccagttgactgactacactcctcgaag 1166
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
DB 1226 aaagtggagtgacatgaaagacacagacagagc---tgcacagcgccactatgtaactg 1282
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
QY 1167 cgtgtgttattatgagctgcagaatgaaggtggtctatcaagaatgacattggttgc 1226
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
DB 1283 tattgtatcatgaaacagagacttaaggtgctgttataatcaccaagcctgtttgc 1342
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
QY 1227 taataaattatcttaaggtggaacctatcttgatataaatgcttggcaagaagcttga 1286
   || || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1343 taacaagtttgagcttaatactacccttactcgttgaaatgactgaactgagagctacg 1402
   || || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
QY 1287 agaa 1290
   |||
DB 1403 cgaa 1406

```

```

RESULT 12
AAV30006
ID AAV30006 standard; cDNA; 1807 BP.
XX
AC AAV30006;
XX
DI 11-AUG-1998 (first entry)
XX
DE Full length cDNA sequence of human IGT.
XX
KW Human; beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme;
KW IGT; core 2-beta-1,1-N-acetylglucosaminyltransferase; C2GNT;
KW treatment; disorder; under expression; type-II hypersensitivity reaction;
KW neonatal haemolytic disease; autoimmune haemolytic anaemia;
KW thrombocytopenia; ds.
XX
OS Homo sapiens.

```

```

XX US5766910-A.
XX 16-JUN-1998.
XX 07-JUN-1995; 95US-0488135.
XX 09-SEP-1993; 93US-0118906.
XX (LJOL-) LA JOLLA CANCER RES FOUND.
XX Blerhuizen MFA, Fukuda M;
XX WPI; 1998-361697/31.
XX P-PSDB; AAW56628.
XX
XX New nucleic acid sequences and their complementary sequences -
XX useful for producing fragment of recombinant human I-branching
XX -1,6-N-acetyl-glucosaminyl-transferase polypeptide
XX
XX Claim 2; Columns 29-32; 31pp; English.
XX
XX The present sequence encodes human
XX beta-1,6-N-acetylglucosaminyltransferase, the I-branching enzyme (IGNT).
XX The protein sequence has a type II transmembrane topology that consists
XX of a short amino terminal cytoplasmic sequence, a signal-anchor sequence
XX followed by a short stem region and a large carboxyl terminal catalytic
XX domain. The protein shows some homology to core
XX 2-beta-1,1-N-acetylglucosaminyltransferase (G2GNT). The nucleic acid
XX sequence is used for producing IGNT polypeptides, which can be used
XX to treat disorders arising from under expression of IGNT, e.g. increased
XX susceptibility to type II hypersensitivity reactions such as neonatal
XX haemolytic disease, autoimmune haemolytic anaemia and thrombocytopenia.
XX
XX Sequence 1807 BP; 511 A; 401 C; 398 G; 497 T; 0 other;

```

```

Query Match 13.6%; Score 185.2; DB 19; Length 1807;
Best Local Similarity 52.1%; Pred NO.4.9e-42;
Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```

```

QY 327 ttgtgacattatcagaactgaagggtatgctcaaaaagctgtctcaaggaggagaa 386
DB 467 ttgcaagaaactgtgacccagagccacatcacagcccttattcgaagaagc 526
QY 387 aagttcccaatagctattcttggtgtccacaagaagcaatttggtgaaggtc 446
DB 527 tgacttccctgacatatataatggtcattccacatcacttgaacaccttgcagaagc 586
QY 447 tatcatgctatatacacacagacacaaatlattactgcatcatcatatgctaagagc 506
DB 587 cttagggctattacatgcccccaaatatctactgtgttcattgtgaagaagaac 646
QY 507 tgatcactcaaaattgccaatgaacaattagcgaagtgtctcccaattttcattgc 566
DB 647 aactgaattcaagaatgcgtgaaggcaacttaagaactgtctcccaacgctttctgyc 706
QY 567 ttccaattagaagctgtgaaatagccacattccagaactccagagctgattaaattg 626
DB 707 ttccaagatgaaccgctgtctatgagaggtctccagagctccagctgaccgaactg 766
QY 627 ctgtgagactcttgaagcttcaaccagtggaataatgtaatacaacttggtggca 686
DB 767 catagagatcttctgctcctcgagagctcatggaagtaagttatcaacacctgtggca 826
QY 687 agatttccctgaagcaaatgttgtaattgtgtcagagttgaaanaaacatcaatgagc 746
DB 827 agacttccctgaagaacacacgaagaatagttcagatcgtgaagaagatttaagtaa 886
QY 747 aaatatgttgagaagcgtgaacccccaacacgaatgaattggaagaattcaccatca 806
DB 887 aaatatacccccagggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 946

```

```

QY 807 tgaactaagcgggtgcttargatatgtgaagctaccacaagaacacaatccca 866
DB 947 ccaagagcaccttgggca-----aagagcttcccatgtgataaagaacagcgttaa 1000
QY 867 ggaagcaccaccacaacatccagatattgtgacgtgtcttattgttttaagta 926
DB 1001 accgctccccccacaataactcaaatctacttggctgtgctgtgtgtgtctaaag 1060
QY 927 agcatltgttaatatatttcaacaacatccatcgttcaagaacttttgcctgtctaa 986
DB 1061 agagttggcaacttgttctgcatgacccacagcgctgtgattgtctcagtggtcaa 1120
QY 987 agacacatctcctcgtgaagacacttgggtgctacactgtatcgtgttcaggaataac 1046
DB 1121 ggaacttccagctcctgagatcattctgtgtgaactcaataagattccagtggtcc 1180
QY 1047 tgggagagattccagatcagccagagatgtgtcatctgcagagtaagactgcctgt 1106
DB 1181 tgccttatgccaatgcatctctgactg-----gaaacctcagagctat 1225
QY 1107 caagtgaaattactatgaagcttttctatccagttgactgatatccactcagag 1166
DB 1226 aagtggaagtgacaatggaagacagacagagagc---tgccagggcacctatgtacatg 1282
QY 1167 cgtgttatattgaagctgcagaattgaagtggtctatcaagaatggacattgttgc 1226
DB 1283 tattgtatcatgaaacagagactaaagtgctgtgttaattcacaagcgtgttgc 1342
QY 1227 taataattgtattcgaagtgagcctctctgtatgaatgcttgcagagaagaacttga 1286
DB 1343 taacaagttgagcttaataactacccctactgtgaaatgctcagaactgaagcatcg 1402
QY 1287 agaa 1290
DB 1403 cgaa 1406

```

RESULT 13

```

AAV16000
ID AAV16000 standard; CDNA; 1807 BP.

```

```

XX AAV16000;
XX AC
XX DT 02-JUN-1998 (first entry)
XX DE Human beta-1,6-N-acetylglucosaminyltransferase (IGNT) encoding cDNA.
XX KW beta-1,6-N-acetylglucosaminyltransferase; IGNT; tumour cell adhesion;
XX antigen binding fragment; antibody; I-branching enzyme; human; ds.
XX OS Homo sapiens.
XX FH Key
XX FT CDS Location/Qualifiers
XX FT 255..1457
XX FT /*tag= a
XX FT /*product= "IGNT enzyme"
XX
XX US5731420-A.
XX PD 24-MAR-1998.
XX
XX PF 07-JUN-1995; 95US-0486196.
XX
XX PR 09-SEP-1993; 93US-0118906.
XX PR 07-JUN-1995; 95US-0486196.
XX
XX PA (LJOL-) LA JOLLA CANCER RES FOUND.
XX PI Blerhuizen MFA, Fukuda M;
XX DR WPI; 1998-216555/19.
XX DR P-PSDB; AAW47184.
XX

```


	Query March	Best Local Similarity	Matches 183;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	60	tttaaccctatgagctgtctctctcttctgttaagaactcttaaatgtgaagacactcttcgca	13.4%;	Score 183;	DB 21;	Length 186;	
Db	1	tttaaccctatgagctgtctctctcttctgttaagaactcttaaatgtgaagacactcttcgca	100.0%;	Pred. No. 7.5e-42;			
QY	120	aaaagacgttacttggtgtgtgtactccttaagtaaccctgcgctcttcttgaagaacagata	17.9%;	Score 183;	DB 21;	Length 186;	
Db	61	aaaagacgttacttggtgtgtgtactccttaagtaaccctgcgctcttcttgaagaacagata	100.0%;	Pred. No. 7.5e-42;			
QY	180	cactcatgtttaagatgaagtcaggaatgaagtaagtaactgtctcgggtatctctatgaacagga	23.8%;	Score 183;	DB 21;	Length 186;	
Db	121	cactcatgtttaagatgaagtcaggaatgaagtaagtaactgtctcgggtatctctatgaacagga	100.0%;	Pred. No. 7.5e-42;			
QY	240	gcc 242	100.0%;	Score 183;	DB 21;	Length 186;	
Db	181	gcc 183	100.0%;	Pred. No. 7.5e-42;			
RESULT	15	AA061559					
ID	AA061559	standard; cDNA; 2105 BP.					
AC	AA061559;						
DT	07-OCT-1994	(first entry)					
DE	CDNA sequence of human core 2 beta 1-6						
XX	N-acetylglucosaminyltransferase (C2GNT or core 1-6 AGT).						
XX	C2GNT; 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase;						
KM	O-glycan; ss.						
OS	Homo sapiens.						
Key	Location/Qualifiers						
FFH	220..1504						
FFH	/*tag= a						
FFH	polyA_signal	1913..1918					
FFH	/*tag= b						
FFH	misc_signal	248..314					
FFH	/*tag= c						
FFH	/label= signal/membrane-anchoring domain						
PN	EP590747-A.						
PD	06-APR-1994.						
XX	29-SEP-1993;	93EP-0250266.					
XX	01-OCT-1992;	92US-0955041.					
XX	(LJOL-) LA JOLLA CANCER RES FOUND.						
PI	Bierhuizen MFA, Fukuda M;						
XX	WPI; 1994-11195/14.						
XX	P-PSDB; AAR51386.						
XX	New beta 1-6 N-acetylglucosaminyl transferase and acceptor - are						
PT	used for the study of the effect of variant O-glycan(s) on						
XX	cell-cell interactions, partic. in cancers						
PS	Disclosure; Page 20-22; 34pp; English.						
CC	C2GNT, or an active fragment thereof, catalyses the formation of						
CC	critical branches in O-glycans. cDNAs encoding various						
CC	glycosyltransferases can be isolated by transient expression of cDNA						
CC	in recipient cells, e.g. COS-1. COS-1 cells were transfected						
CC	with a cDNA library, pcDSR alpha-2FL, constructed from poly(A)+ RNA						

of activated T lymphocytes which express the c2GnT. Transfected CC cells were selected using Mab T305¹ which identifies a CC hexaasaccharide on leukosialin. Leukosialin CD43 is an acceptor molecule for c2GnT activity. A plasmid, pcDSR alpha-Leu, which directed expression of the T305 antigen was identified. The CC cDNA insert was isolated and sequenced. The cDNA encoded the acceptor molecule leukosialin CD43.

Sequence 2105 BP; 614 A; 402 C; 500 G; 589 T; 0 other;

Query Match 12.9%; Score 175.2; DB 15; Length 2105;
Best Local Similarity 51.9%; Pred. M.34e-39;
Matches 517; Conservative 0; Mismatches 443; Indels 36; Gaps 4

```
QY 300 tgaatgattgttgagcaatgacacagcagatggtatcatattacacgacttaagaagttatgc 359
Db 489 tgaagcattataaacaatgacacagcagctgtctctcttcatcagaagaccaaatat 548
QY 360 tcaaaagctctgtctcaaaaggaggaaaaaagcttcccaatagccttcttggttgtcca 419
Db 549 tgtagaaccccttaagaagaagcgaggttcccaaaagaatcttcataagttgtcca 608
QY 420 caaagatgcataatgattggtgaagaagcttccatgcatatatacaacacagacaattta 479
Db 609 tcaaaagattgaatcgtttgacagcgtgtgtggtgcacatctatccctcagaatttcta 668
QY 480 ctgcgtccattatgtaatcgaagaagcaccgtatcccttcaaaagttgcacatgaacaatttagc 539
Db 669 ttgcttctctgtgcagacacaacaacccagagatcttcatttaagctgcagttatgtgcacgc 728
QY 540 taagtgctctccaatatttcatattgcttccaatataagaagcgctgtggaatatgcccacat 599
Db 729 ttccgttttaagtatcttctgttgccagccgattgttgagagttgtgtttatgcacgtg 788
QY 600 ttccgagctcccgagctgatttaattgctctgtcggagccttcgaaagtcttcaatccagtg 659
Db 789 gagcgagttccagctcgacccccaactgtcagtaagaagctctctgcacatgagtaacaacgt 848
QY 660 gaataatgttacaacttctgttgaggcaaaatttccctcgtgaagtcacaatttgaattggt 719
Db 849 gaagtcttgataaactcttctgtgtatgtatcttcccatlaaacaacactagaattgt 908
QY 720 gtccagattgaaaaaactcaatgatgagccaatatgtttgtagagcggtgtaaacccccaag 779
Db 909 caggaaagctcaagttgttataatggaagaaaaacaactcgtgaacgggaagagatgccacca 968
QY 780 taaatggaagaatttcaacttaccatcatgaactatagaacggtgtccctatgaatatgtga 839
Db 969 taagaagaagaagttggaagaagcgttatga-----gltcgttaatggaagaagctgac 1019
QY 840 gttaccaataaagacaacaatctccagaagaagcacccccacataacatccagatatgtt 899
Db 1020 -----aaacacgggaactgtcaaaatgtcttccccaactcgaacacccctcttcc 1070
QY 900 tggcagtgcttatttgttttaagtcacaagcatttgttaaatatatttcaacaactccat 959
Db 1071 tggcagtgcttactctgtgtcatgttagtgagatgtggtgtatgtactacaatatgaaa 1130
QY 960 cgttcaagacttttggcctgtctaaagaacacatctctccgtcagaagaacatttgggc 1019
Db 1131 aatccaaaagttgattgagatgggcacaagaacacatcaacgcttgatgaattctcgggc 1190
QY 1020 taactgtatcgggttccagaagaataccctggggagat--ttcagaatcagccagaagt 1076
Db 1191 caccatccaagaagattccgtaagctcccggtccatccctgcacagacataagatatact 1250
QY 1077 gactgtatcgcagagaagaagaccgcgcttgcagttgaattactatgaaggcctttt--- 1133
Db 1251 acttgacatgcagacagcgttgcaggttttcaagttgcagctactttgaagggtgatgttcc 1310
QY 1134 -----ctatccagattgtatcgtgatactcaacttcgaagcggtgttatatg 1181
```

Fri May 3 10:57:52 2002

us-09-645-192-1.rng

Page 15

Db 1311 caaggtgctccctaccgccctgcgtgagtcacatgtgcctcaagtgtgcaatttcgg 1370
QY 1182 agctgcagaatttaaggtgcttacaagaatggacattggttgcataataattgattc 1241
Db 1371 agctggtgacttgaactggaatgctgcgcaaacaccacttggttgcataaagttagcgt 1430
QY 1242 taaggtgaccatacttgaattaaatgcttggcaga 1277
Db 1431 ggaatgtgacctcttgcacatccagtgttggatga 1466

Search completed: May 2, 2002, 11:37:34
Job time: 7474 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 08:18:30 ; Search time 1511.47 Seconds
(without alignments)
9683.123 Million cell updates/sec

Title: US-09-645-192-1
Perfect score: 1362
Sequence: 1 atgaagatttcacatgtta.....atctacacacacatcarga 1362

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estlin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gb_est1: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rtd: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	437.8	32.1	821	11	BG434801
2	408.4	30.0	410	13	AC005888
3	393.2	28.9	612	13	AC0900348
4	350.8	25.8	590	13	A2603057
5	289	21.2	292	13	AC091453
6	253	18.6	908	13	CNS02A46
7	221	16.2	973	13	CNS058A
8	187	13.7	1367	12	AK019924
9	174.2	12.8	1862	12	AK008234
10	170.8	12.5	860	12	AL576150
11	163.2	12.0	970	13	CNS04UIM
12	157.6	11.6	756	13	CNS01UWZ

C 13	148.8	10.9	526	10	BE005007	BE005007 MRO-BN011
C 14	148.8	10.9	612	10	BE292814	BE292814 601105179
C 15	142	10.4	873	11	BC185546	BC185546 RST4455 A
C 16	139.4	10.2	567	13	A2388491	A2388491 TM0148603
C 17	136.6	10.0	798	11	BE220853	BE220853 RST40650
C 18	135	9.9	774	11	BE204638	BE204638 RST24051
C 19	133	9.8	572	10	BE234997	BE234997 142220 MA
C 20	133	9.8	655	10	AM842622	AM842622 MR2-CN003
C 21	129.8	9.4	1050	11	A1722764	A1722764 fc31c11.Y
C 22	127.8	9.2	869	10	BE867668	BE867668 601443206
C 23	124.8	9.2	869	10	AL555400	AL555400 AL555400
C 24	121.6	8.9	820	11	BC198987	BC198987 RST18281
C 25	118.2	8.7	695	11	BF036111	BF036111 601457742
C 26	116.2	8.5	731	11	BE468641	BE468641 602510243
C 27	115.8	8.5	842	11	BE869192	BE869192 601445191
C 28	113.8	8.4	554	10	AA307800	AA307800 EST178656
C 29	113.8	8.4	671	10	AM233332	AM233332 f129g12.Y
C 30	110.8	8.1	788	10	A1528293	A1528293 u195f11.Y
C 31	108.8	8.0	791	11	BE788203	BE788203 SP-0035_B
C 32	107	7.9	428	13	AQ437509	AQ437509 HS-513_B
C 33	102.4	7.5	506	11	BI337733	BI337733 361252 MA
C 34	102	7.5	348	11	BE863271	BE863271 UI-M-BHO-
C 35	101.2	7.4	482	11	BF416888	BF416888 UI-R-CNO-
C 36	99.6	7.3	542	10	AM388627	AM388627 MR2-ST013
C 37	97	7.1	633	10	A1133042	A1133042 HA1630.Hu
C 38	96.6	7.1	398	13	A2149281	A2149281 SP-0035_B
C 39	95.8	7.0	622	10	A1747955	A1747955 UI04007.Y
C 40	93.4	6.9	605	13	AL194803	AL194803 SP-1028_B
C 41	92.4	6.8	555	10	BE721246	BE721246 188355 MA
C 42	90.6	6.7	443	11	BE755013	BE755013 QVO-CT058
C 43	88.6	6.5	489	10	BE721377	BE721377 188533 MA
C 44	88.2	6.5	438	10	AM842590	AM842590 MR2-CN003
C 45	87	6.4	543	11	BE077138	BE077138 225944 MA

ALIGNMENTS

RESULT 1
BG434801
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1
BG434801 821 bp mRNA
602507322F1 NIH_MGC_79 Homo sapiens CDNA clone IMAGE:4604659 5',
BG434801
BG434801.1 GI:13341307
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 821)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov> row: k column: 06
Plate: L10M1340
High quality sequence stop: 709.
Location/Qualifiers
1..821
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4604659"
/clone_lib="NIH_MGC_79"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: Placenta; Vector: pNR-LIB (Clontech);

FEATURES

source

Site_1: 5'fl (ggcgccctggccg); Site_2: 5'fl (ggccattatggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCCGAGCGGCGGACATG-3' (30)BN-3'
(where B = A, C, G, or T). Average
insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library."

BASE COUNT 241 a 159 c 163 g 256 t 2 others
ORIGIN

Query Match 32.1%; Score 437.8; DB 11; Length 821;
Best Local Similarity 99.5%; Pred. No. 3e-102;
Matches 439; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 922 agtcaagcattgtttaataatattttcaaacatccatccgttcaagacttttggcctgg 981
DB 1 AGTCAAGCATTGTGTTAAATATTTCACAACTCCATCATTCAGACTTTTGGCCTGG 60
QY 982 tccaagacacatactctctctgatagagcatttgggctacattgacgggttccaga 1041
DB 61 TCTAAAGACACACTACTCTCTGATGAGCACTTTGGGCTACTGATTGGGTTCCAGGA 120
QY 1042 ataccggggagagttccagatccagccagagatgtctgcatctcagagtaagctcgc 1101
DB 121 ATACCTGGGGAGATTTCAGATCCAGAGATGTCTGATCTCCAGATGAGTACGCTGC 180
QY 1102 ctgtcaagtgaattactatgaaggctttttctatcccaagttgactgatacctt 1161
DB 181 CTGTCAAGTGAATTAATGATGAGAGGCTTTTCTATCCAGTTGACTGATCTACCTT 240
QY 1162 cgaagcgtgtgtttatgagcgtcgaagtaagttggtctataaagatggaattgg 1221
DB 241 CGAAGCCTGTGATTATGAGCTGCAAGATTGAGTGGCTTATCAAAATGAGATATGG 300
QY 1222 ttgtctaataattgattcttaaggtggaacctatcttattaaatgcttggcagaaga 1281
DB 301 TTTCCTAATTAATTTGATTCTAAGTGAAGCCATCTGATTAATGCTTGGCAGAAAG 360
QY 1282 ctggaagaacagcagagagatgatacattttgcccctcgaagaagtattttgtatga 1341
DB 361 CTGGAAGACAGCAGAGAGACTGATCTTGTCTCAGAAAAGTTATTATGATGAGA 420
QY 1342 aatctactaccacatcatga 1362
DB 421 AATCTCACTACCAATCATGA 441

RESULT 2
LOCUS A0005888/c 410 bp DNA 27-JUN-1998
DEFINITION CIT-HSP-2288B17.TF CIT-HSP Homo sapiens genomic clone 2288B17, DNA
sequence.
ACCESSION A0005888 GSS
VERSION A0005888.1 GI:3083333
KEYWORDS GSS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 410)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other GSS: CIT-HSP-2288B17.TF
Contact: Mark Adams
Department of Eukaryotic Genomics

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mddams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M3-21;
Class: BAC ends

FEATURES
source Location/Qualifiers
1..410
/organism="Homo sapiens"
/db_xref="GDB:7149764"
/db_xref="taxon:9606"
/clone="2288B17"
/clone_1lb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 135 a 87 c 81 g 107 t
ORIGIN

Query Match 30.0%; Score 408.4; DB 13; Length 410;
Best Local Similarity 99.8%; Pred. No. 1e-94; Indels 0; Gaps 0;
Matches 409; Conservative 0; Mismatches 1; Indels 0;

QY 869 aagacccccccataacatcagatatgttgggaagtctatttgttttaagtcag 928
DB 410 AACACACCCCCCATTAATCATTCAGATATATGTTGGCAGTCTTATTTGTTTAACTCAG 351
QY 929 catitgttaataattttcaaacatccatgcttcaagaacttttgcggtctaaag 988
DB 350 CATTTGTTAATTAATTTTCAACAACTCCATGCTTCAACACTTTTTCCTGGCTTAAG 291
QY 989 acaatactctctctgtagagacatttgggtctacacttgatcttgggttccagaatactg 1048
DB 290 ACACATCTCTCTGATGAGCACTTTGGCTGACTGATGAGGTTCCAGAAATACCTG 231
QY 1049 gggagatttcagatcagaccagagatgtctgatactgagagtaagatcgcctgtca 1108
DB 230 GGAAGATTTCAGATCAGCCAGAGATGTCTGATCTGCAAGTAAGATCCCTTGCA 171
QY 1109 agtgaattactatgaaggctttctatcccaagttgtaacgtgataccttgaagag 1168
DB 170 AGTGAATTAATTAATGAAGGCTTTTCTATCCAGTTGATCTGATCTGACCTTCAGAGCG 111
QY 1169 tgttattttgagcgtcgaagtaagttggttatacaagaatgacattgttgcata 1228
DB 110 TGTGTATTTTGAAGCTGCAAGATTGAAGTGGCTTATCAAAATGAGACATTGGTTGCTA 51
QY 1229 ataaattgattcctaaggtggaacctatcttgaataatgcttggcagaa 1278
DB 50 AATAATTGATTCTAAGGTGAGCCCTATCTGATTAAATGCTTGGCAGAA 1

RESULT 3
LOCUS A0900348/c 612 bp DNA 10-NOV-1999
DEFINITION HS_2175_B1.A12.T7 CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-2175 Col-23 Row-B, DNA sequence.
ACCESSION A0900348 GSS
VERSION A0900348.1 GI:6356538
KEYWORDS GSS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 612)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

TITLE
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE
99380589
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2175 Row: B Column: 23
Seq primer: T7
Class: BAC ends
High quality sequence stop: 612.
Location/Qualifiers

FEATURES
source
1. 612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIR Approved Human Genomic Sperm Library D"
/sex="Male"
/note="Organ: Sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 180 a 141 c 121 g 165 t 5 others
ORIGIN

Query Match 28.3%; Score 393.2; DB 13; Length 612;
Best Local Similarity 91.6%; Pred. No. 9e-91;
Matches 427; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 839 agctaccaataaggacacacatcccaaggagagaccccccaataacattcatttg 898
Db 610 AGCTCCAAATAGAGTCAAGCTCTCCAGAGACACCGCTCATACATTCAGATATG 591
QY 899 ttggc-agtgcattattgtttaagcagcattgttaataataatttcacaactcc 957
Db 550 TTGGCAGTGTATTTTGTTCAGTCAGCATGTGTTAAATATTTTCAACACTCC 491
QY 958 atggttaagactttttgcttgcttgctcaagaacacatactccctgatgagcatttg 1017
Db 490 ATGCTTCAAACTTTGTGGTGTGCTAGAGACACATCTCATGTGAGCACTTGTG 431
QY 1018 gctaccttgattcggttcgaagataacccgtggagattccagatcagccagagtgc 1077
Db 430 GTTACCTTGATTCGGGTTCAGAAATACCTGGGAGATTTCCAGATCAGCCAGATG 371
QY 1078 tctgattcgaagtaagactgccttgtaagtggaattactataagcttttctat 1137
Db 370 TCTGATCTGAGAGATGAGATCCCTTGCTCACTGCAATTAATTAAGAGCTTTTCTAT 311
QY 1138 ccaagtgtactgattcacccttcgaagcgtgtgattatgagagctgcagattaag 1197
Db 310 CCCAGTGTACTGATTCACCTTCGAAAGCTGTGTAATTTATGAGACTGACGATTAAG 251
QY 1198 tggccttacaagaatgacattggtttgctaataaatttgatctaaagtgagacctatc 1257
Db 250 TGGCTTATCAAAAGATGACATGCTTGTGTAATTAATTTGATCTAAAGGAGACATATC 191
QY 1258 ttgattaagcttcgcagaagaagcttgtaagaacacagagagact 1303
Db 190 TTGATTTATATGCTTGGAGAAAGCTTGCGTGCAGAGTGCCT 145

RESULT 4
AZ603057 590 bp DNA
LOCUS 1M0422E09F Mouse 10kb plasmid UUCG1M library Mus musculus genomic
DEFINITION clone UUCG1M0422E09 F, DNA sequence.

ACCESSION
AZ603057 GI:11725247
VERSION
GSS
KEYWORDS
house mouse.
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 590)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
JOURNAL
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
64112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0422 Row: E Column: 09
Seq primer: CGTGTAAAGACAGCGCCAGT
Class: Plasmid ends
High quality sequence stop: 590.
Location/Qualifiers

FEATURES
source
1. 590
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUCG1M0422E09"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: pMD29v; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (91473211419b1AFL29072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance.

BASE COUNT 157 a 147 c 138 g 147 t 1 others
ORIGIN

Query Match 25.8%; Score 350.8; DB 13; Length 590;
Best Local Similarity 79.6%; Pred. No. 7.6e-80;
Matches 440; Conservative 0; Mismatches 108; Indels 5; Gaps 2;

QY 1 atgaagatatacaaatgtattttaaatacctacagagaagtttcatcctgtt 60
Db 38 ATGAGATATTCAGATGTTGCTTAAATACACTCCAGCAGAACTTCTCATCTC 97
QY 61 ttaaccctatgctctctcttcttgaagcttctaattg---agaagcatttcgcg 117
Db 98 TTAACCTGTGCGCTGTTCTCTTGTGAAGCTCTTAATGAGGAGGCTCTTCTCC 157
QY 118 caaaagacattacttggttgagtaaccttaagtaacctgccttttgtaagaacaga 177

[illegible]

RESULT	6
CNS02AAG/C	GSS 12-MAY-2000
LOCUS	Tetraodon nigroviridis genome survey sequence PUC-Crl end of clone 25IC11 of library G from Tetraodon nigroviridis , genomic survey sequence.
ACCESSION	CNS02AAG 908 bp DNA
VERSION	AL186089.1 GI:7826193
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis . Tetraodon nigroviridis . Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Notolestei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 908) Reest,Crollius,H., Jalllon,O., Dasilva,C., Fizeser,C., Fisher,C., Bourneau,L., Billaut,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
TITLE	Unpublished
JOURNAL	2 (bases 1 to 908)
REFERENCE	Reest,Crollius,H., Jalllon,O., Dasilva,C., Bourneau,L., Fisher,C., Bernot,A., Fizeser,C., Wincker,P., Brottler,P., Quetier,F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
AUTHORS	Unpublished
TITLE	3 (bases 1 to 908) Genoscope.
JOURNAL	Direct Submission
COMMENT	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .
FEATURES	Location/Qualifiers
Source	I..908

```
Location/Qualifiers
1.: 908
/organism="Tetrahodon nigrivittatus"
/db_xref="taxon:99883"
/cclone="251C11"
/cclone-1b="G"
/note="Microscope sequence ID : C0AG251AB06SP1-end
CUC-01"
```


Db	108	AGCGGGTGAAGGCCGACCTCACTGCTGTCCGAGACTTCTGAGAGGTGCAAGTCAAGTGG	149
Qy	661	aaataatgtatccaacttggtgtggcgaagatattcccttgtaagttcaatttgatgtg	720
Db	148	AAGTACCTCATCAACCTCTCGCGGCCAGGATTTTCCCTCAGGTCCAAACATCGAGCTGGT	89
Qy	721	tcgaagtgtgaaaacacccaatgagcaaatatgttggagacggtgtaaccccccaaacagt	780
Db	88	TCGAGCTGGAAGAAACTAAAGCGGGCTAATATGTTAGAGAGCGCGCACCCACAGACTAT	29
Qy	761	aaatgtgaaagattcacttac	801
Db	28	AAGAGCGAGAGATTCACTTTC	8
RESULT	8		
LOCUS	AK019924		
DEFINITION	AK019924	1367 bp	mRNA
ACCESSION	AK019924		05-JUL-2001
VERSION	AK019924.1	GI:12860326	
KEYWORDS	cap-trapper.		
SOURCE	Mus musculus (strain: C57BL/6J) adult male pituitary gland cDNA to mRNA, clone: lib:RIKEN full-length enriched mouse cDNA library clone:5330430K10.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
METHODS	Methods in enzymology. 303, 19-44 (1999)		
PUBMED	99279253		
REFERENCE	10349636		
AUTHORS	2 (bases 1 to 1367)		
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
METHODS	Genome research. 10 (10), 1617-1630 (2000)		
PUBMED	20499374		
REFERENCE	11042159		
AUTHORS	3 (bases 1 to 1367)		
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, Y., Nishi, K., Katsuna, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, N., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, T., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, K., Yoneoka, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsui, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed capillary sequencer		
METHODS	Genome research. 10 (11), 1757-1771 (2000)		
PUBMED	20350913		
REFERENCE	11076861		
AUTHORS	4 (bases 1 to 1367)		
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
JOURNAL	Functional annotation of a full-length mouse cDNA collection		
PUBMED	Nature 409, 665-690 (2001)		
REFERENCE	5 (bases 1 to 1367)		
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Kono, H., Arai, A., Aikawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiroaki, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Okazaki, Y., Okubo, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sekai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A.,		

TITLE
JOURNAL

Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel.:81-45-503-9216,
Fax:81-45-503-9216)

COMMENT
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to Prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAAGAAATCCCAACAGCTTGTATTTCCTTTTNN 3'], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 3.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAAGAAATCCTCGAATAATTAATACCACCCCCCCC 3']. cDNA was cleaved with BamHl and XhoI. Vector: a modified plusscript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DHIOB.

FEATURES
source
Location/Qualifiers
1..1367
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1912211"
/db_xref="MGI:MGI:1925531"
/clone="531045OKR10"
/sex="male"
/tissue_type="pituitary gland"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
235..1305
/note="putative"
/codon_start=1
/protein_id="BAB31918.1"
/db_xref="gi:12860327"
Translation="MSLRGVAVSAISLVIFVVFYSQLSPLNYOOLNSSSERTSY TICDGLONHFFFTGGDSPHLRLSPQRISHYTSLSEEAFLPAYIMVIHKDFDELFRAIVMCNVVCVHDASKADTFKEKRVQLLCPFAVLASVEOVYVIGESRDALNMKLDAVKVMKVYLNTGDGFPLAKTKELINHLKRGRKNIPGVVPAYIVTRKYVDHERKKDGEYMHTKIILTPPHOLIIEGTAVVALFDVFNI NDERIALTEMSKDPSDEHPFVTNLNPVGSGMPNAWMTGLRAVKKMDMAKKGCHGVKSQTLPVADPROI"

CDS
BASE COUNT 347 A 368 C 336 G 316 T

ORIGIN

Query Match 13.7% Score 187; DB 12; Length 1367;
Best Local Similarity 55.0%; Pred. No. 1.4e-37;

MATCHES
Matches 392; Conservative 0; Mismatches 315; Indels 6; Gaps 1

Dn 637 AAAGAACGGTGCGGCAGACTTAACTGCTTTGCCCATTCCTTCGGCCTTAGAGG 696

OY	577	gagctgtggaaatcttgcacattccaggactccaagctcatatgaatttcctgtcggac	636
Db	697	GAGCAGAGTGTCTATGTTGGCTTCCTCTCGGCTCCAGGCTATCTGAACCTCATGAAGAAT	756
OY	637	ctcttgaaagtcttaatccagttgaaaatatgtttaacaatttgttgyggcaagaattcc	696
Db	757	CTGTGGGCCCTCAAGTGTCCTCCCTGGAAATATGTCTCTCAACACTGGCGGGCAGACTTCCT	816
OY	697	ctgaagtcataatttgaaattgtgtcgagtttgaagaaaaactcaatggagaanaatgtg	756
Db	817	TTGAAAACCACAAGGAATATTATAACCATCTGAAAAGATTTAAGGGAAAAACATCACT	876
OY	757	gagacgvtgnaaccgcccaaacagtaaatitggaagaattcctaaccaicaigaactaga	816
Db	877	CCGGAGAGTGCCTGCCCTCGCTACATGTTGTACGAGCTAATATGTATGACACGAAACGC	936
OY	817	cgggtgccttttgaaattgtgaagctcacacaataagaagcaaacatctccaagagaacc	876
Db	937	AAGGACAAAGAGGATATTATTATC-----ATTAACCAATATTATTGAAGACTCCACCT	990
OY	877	ccccataacattccagatatttgttggcagtgcttatcttttttaagcaagcatgtgt	936
Db	991	CCACACCAATGANCATCTACTTCTGGCACAGCCACTGCGCCCTACCCGGACATTGTCT	1050
OY	937	aatatattttcaacaacctcatctgttcaagaacttttttcctgttctaagaacacatac	996
Db	1051	AACCTTATCCGGAATGACGAAAGGGCCATGCTCTCTTAAGTGGTCTTAAGAATACCTAT	1110
OY	997	tctcctgatgagcaacttttggcctcaccttgatctcggcttccaagtaactctg	1049
Db	1111	AGCCCTGATGAACATTTTGGGCTGACACTCAATAGSATTCOCAGAGTCCCTCG	1163
RESULT	9		
AKO08234			
LOCUS	AKO08234	1862 bp	mRNA
DEFINITION	Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010013H22, full insert sequence.		
ACCESSION	AKO08234		
VERSION	AKO08234.1	GI:12842295	
KEYWORDS	CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to mRNA, clone:11b;RIKEN full-length enriched mouse cDNA library clone:2010013H22.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1862)		
TITLE	Carninci,P., Shibata,Y., Hayatsu,N., Sugihara,Y., Shinata,K., Itoh,M., Konno,H., Okazaki,Y., Mizumatsu,M. and Hayashizaki,Y.		
JOURNAL	High-efficiency full-length cDNA cloning Methods in enzymology. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2 (bases 1 to 1862)		
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugihara,Y., Shinata,K., Itoh,M., Konno,H., Okazaki,Y., Mizumatsu,M. and Hayashizaki,Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome research. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3 (bases 1 to 1862)		
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Kikuma,Y., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujikawa,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Mizumatsu,M., Inoue,K., Kiria,A. and Hayashizaki,Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer		
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)		

Query Match	BASE COUNT	ORIGIN	FEATURES	COMMENT	TITLE	JOURNAL	AUTHORS	REFERENCE	PUBMED	MEDLINE
12.8%	Score 174.2; DB 12; Length 1862;									
	484 a 463 c 478 g 437 t		<p>location/Qualifiers</p> <p>1..1862</p> <p>/organism="Mus musculus"</p> <p>/strain="C57BL/6J"</p> <p>/db_xref="taxon:10090"</p> <p>/db_xref="MGD:MGI:1904808"</p> <p>/db_xref="MGD:MGI:1919337"</p> <p>/clone="2010013H22"</p> <p>/sex="male"</p> <p>/tissue-type="small intestine"</p> <p>/clone_lib="RIKEN full-length enriched mouse cDNA library"</p> <p>/dev-stage="adult"</p> <p>240..1436</p> <p>/note="putative"</p> <p>/codon_start=1</p> <p>/protein_id="BAB25548.1"</p> <p>/db_xref="GI:12842296"</p> <p>/translation="MDIDSEFQYCRDLTYKTKLRPAKSSINGVINGEOKAAYQ</p> <p>ALNNLEIKKKQCLFTFADYLRMTADEDEHKTRKFLQYPLSKREASFPPIAYSNVBE</p> <p>KIENEFELRLAVITPQVYCHMDQSEPFKQAVIVSCFPAVFLASKVAVYVAS</p> <p>WSRQADLNCEMEDLLQSPVPMKYILNCGDFPKTIAEWAKALIKDQSKSEVYP</p> <p>PPHKSFMKYHYEVTDLHMTSKRRTPPNLITFNENAYVASRDIIEVFNSSKAR</p> <p>QLIEWVDYSPDEHLNATLQRAQSMQSGDPLHKEFMSDMRAIPLRTKYDHEGDIE</p> <p>NGAYTSCGICHRAYCVYSGDLDHILNHHLLANKFDEKVDNVNQCIEEYLRHRA</p> <p>ITYGEL"</p>	<p>CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGAGACGATGATTAATTAATGATCCGCCCCCCC 3']. cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGGAGATTCAGTTAATTAATGATCCGCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.</p>	<p>Functional annotation of a full-length mouse cDNA collection</p> <p>Nature 409, 685-690 (2001)</p> <p>5 (bases 1 to 1862)</p> <p>Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carinici, P., Fukuda, S., Fukunishi, Y., Fununo, M., Hasegawa, T., Harada, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Iwano, M., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, K., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tezima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Yamatsuo, M. and Hayashizaki, Y.</p>	<p>Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Saito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)</p> <p>Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.</p>				

Best Local Similarity 51.5%; Pred. No. 3e-34;
Matches 506; Conservative 0; Mismatches 438; Indels 39; Gaps 3;

QY	316	ataccagcgtattgtgacattatatacagctctaaagagtttctgcacaaagcttgcctca	375
Db	438	ATGACAGCAGACTGTGTAGCAGCTTCAAGACCAAGAGAGAGTTTATATACAGTCCCACTGAGC	497
QY	376	aagagaggagaaaagcttcccaatagcctattctttgtcttcacaaaagtatgcattatg	435
Db	498	AAGGAAGAGGCCAGCTTCCCATTTGGTACTCCATGAGTGAGTGCAGAGAAATTTAGAAC	557
QY	436	gtbaaagggttatccatgcatatatacaacacagcaaatatttaactgcatcattatgat	495
Db	558	TTGCAAGAGTGTGCTGGAGCTGTGTATACCCCTCGAGTATATATGTGTGTCCATATGAT	617
QY	496	cgtaaggcacctgataaccttcaaaagttgcattgaacaaatttagctaaagttcttccat	555
Db	618	CAGAAAGCTTCAAGAACCCTTTAAAGCAGGAGTGCAGGGCCATCTGTATGCTTCCCAAT	677
QY	556	atttactgtctcccaaatatagaaggtgtggaatttgcacacatttccagactccagact	615
Db	678	GTTCTTACTACTAGTATAGTGTGTGTGTAGTGTCTATGCTTCTGTGTCAGGGTGTGAGCT	737
QY	616	gatttaaatgtctgttcgagaccttcgaagttcctaattccagttgnaaatatgtatcaac	675
Db	738	GACCTTAACCTGCATGGAAGACTGTTCTTCAGACCCCGTGCCTCATGMAATACCTCTGAC	797
QY	676	tttgtgtggaagaatttccccctgaagtcgaattttgaattgtgttcaaggttgaaaaaa	735
Db	798	ACCTGTGGGACAGACTTTCCCATCAAAACCAATGCTGAGATGTCTAAGCCCTCAAGCTA	857
QY	736	ctcaatgtgagcaaatatgttctgagacagtggtgaaccccccaagaatattgaaagattc	795
Db	858	TTGAAAGGGCAGAAACATATGAGTGTAGAGGTACCCACCTCCATATAAAATCCCGCTGG	917
QY	796	acttaccatcatgaacttagcagcgtggtccttatgatatatgaaagctcccaataagaca	855
Db	918	AAATATCACTATGTGGTGACA-----GACAACTTCGACATGACC	956
QY	856	aacatctccaaggaagagacccccccaatacatcaatalttgttctggaagtgttatttt	915
Db	957	AGCAAGAGGAAGACGCCGCCACTTAATTAACCTAACCATGTTCACTGGGAGATCCCTACATG	1016
QY	916	gtttaagtcaagatttgtttaatatattttbaaacaatccatcgttcaagacttttt	975
Db	1017	GTGGCTTCTGAGACTTCATTGAACACGTTGTAGTACTCAAAAGCCCGGCACTGATC	1076
QY	976	gacctgtctaaagacacatactctctctgataagacattttggcctacccttgaattcgggt	1035
Db	1077	GAGTGGGMAAAGACACACTTATAGTCCGATGAGACACTTTGGGCCACCCCTCAGCGTGC	1136
QY	1036	ccaagaaataactg---ggagatttccagatacagcccaagatggttcttgatctgcaaggt	1092
Db	1137	TCTGTGAGTCCCTGATPAGATCCCTTGATCGAAAAATTGACATGTGCACATGTAGACCC	1196
QY	1093	aagactcgacctgtgcaagtgtgaattactaactgaagcgtttt-----ccat	1137
Db	1197	ATTGGAGACTAACCCAGGTGTACGACATGAGAGATGAGACAGGGCACCTTAC	1256
QY	1138	ccccgttgatctgatatctcaacttcgaagcgtgtgattatagggagctgtcgaattaaag	1197
Db	1257	ACGTTCTTCTCAGGAATCCACACCGGGCTGTCTGTGTATATGGGTCAAGGGACCTGCAC	1316
QY	1198	tgacctatcaagaatgacatatgtgttgccttaaaaaatttgattcttaaggttgaacctatc	1257
Db	1317	TGGATCTTTCAGAACCATCACTTCACTTGGCCCAACAGTTTGACCAACAAGSTGTGATAT	1376
QY	1258	ttgataataatgtcttgcagaanaa	1260
Db	1377	GTTCTTCAGTGTTTAGGAAGATA	1399

AL576150/c	AL576150	860 bp	mRNA	EST	16-FEB-2001
LOCUS	AL576150	LTT-NEU006_P12	Homo sapiens	CDNA clone	CSDD10721YD06 3
DEFINITION	AL576150	LTT-NEU006_P12	Homo sapiens	CDNA clone	CSDD10721YD06 3
ACCESSION	AL576150	AL576150	AL576150	AL576150	AL576150
VERSION	AL576150.1	AL576150.1	AL576150.1	AL576150.1	AL576150.1
KEYWORDS	EST.	EST.	EST.	EST.	EST.
SOURCE	human.	human.	human.	human.	human.
ORGANISM	Homo sapiens	Homo sapiens	Homo sapiens	Homo sapiens	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 860)	1 (bases 1 to 860)	1 (bases 1 to 860)	1 (bases 1 to 860)	1 (bases 1 to 860)
JOURNAL	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
COMMENT	Full-length CDNA libraries and normalization	Full-length CDNA libraries and normalization	Full-length CDNA libraries and normalization	Full-length CDNA libraries and normalization	Full-length CDNA libraries and normalization
	Unpublished (2001)	Unpublished (2001)	Unpublished (2001)	Unpublished (2001)	Unpublished (2001)
	Contact: Genoscope	Contact: Genoscope	Contact: Genoscope	Contact: Genoscope	Contact: Genoscope
	Genoscope - Centre National de Sequencage	Genoscope - Centre National de Sequencage	Genoscope - Centre National de Sequencage	Genoscope - Centre National de Sequencage	Genoscope - Centre National de Sequencage

BASE COUNT	218 a	158 c	219 g	246 t	19 others
ORIGIN					

Query Match	Similarity	Score	DB	Length
Best Local	54.48%	Pred. No. 26-37		
Matches	371	Conservative	10	Mismatches 289; Indels 12; Gaps 2
Qy	345	tctaaagagttctgcctcaaaagcttgcctcaagagagaaagcttcccaatagccta	404	
Db	767	TGGAACCCACTATGTACACAAACAACTCTGTGAAGAAGGSGTGGTCCCTTTCCTTA	708	
Qy	405	ttccttggttgtccacaaagatgcaattatggttgaagagctatccatgctatacaca	464	
Db	707	CACAGGACCACTCCACAAAAGACTTCGGCACTTTGGAGAGCTTTTAGGCGGATTAAT	648	
Qy	465	ccagacaatatctactgcataccatgatgatcgtlaagagacctgatacttcaaaatgco	524	
Db	647	GCCCCAAATGTCTACTGTGTGTCACCTGGATCGAAGGCGAGCGATCCCTTTAAAGTGC	588	
Qy	525	catgaacaatttagctaagtgctctcccaatatcttcattgtctccaatataagagctgt	584	
Db	587	AGTAAACAGTCACTCAGCTGCTCCCAATGCTTTCTGGCTTCCAAAKAAGTCGGT	528	
Qy	585	ggaataatgccacatttcagacatccagagtgtaattaaattggtgtcgcacctctgaa	644	
Db	527	TGCTATGGGKKAATCCAGDGTCCAMGTGACCTCGAATGCTGCTGAAGACCTTGCC	468	
Qy	645	gtctcaatccagtgaataatagttaaccaactgtgtggcgaagatlttcccttgatgc	704	
Db	467	CTCTGAAGTCCCTGTAAGATTTTCATCAACACCTCGGGCAGAACCTTTCCCTGAAAC	408	
Qy	705	aaatttgatgtgtgcagagttgaaaaaaccaatgagcaatatgtttgagaagcgt	764	
Db	407	CACACAGGAATATGTTCACTACATCTRAAGGATTTTAAAGGAAATAATCACCCCGGATT	348	
Qy	765	gaaaccccccaacagtaaatgtgaaagatctacctt---ccatcatgaacttgaaagcgt	821	

Db 347 GCTCCCTCGACCGCTGTTGGACGACTAATCTCCACCAAGAACTTTAAACCA 288
 QY 822 gcccttaaatgtgaagctacaaataaggaacatccaaaggaacacccccc 881
 Db 287 CAAAATTCCTACGTG-----ATTAACACACAAATTTAAAACTCCTCCBCA 237
 QY 882 taacatcaagatatttctggcagtgcttattgttttaagtaagcatttgaata 941
 Db 236 TGACATGGTATTACTTGTGSCACGGCTGCTCTCAAGGAGACTTGTACTT 177
 QY 942 tatttaacaacatccatcccttaagaacttttgcctgcttcaaaacacatctcc 1001
 Db 176 GGTCTCCAGACAGCAGCTGCACCTTACTCTCTGTGCAAGACACTTACGCC 117
 QY 1002 tgatgaagcatttggagctacc 1023
 Db 116 CGACGAACATTTCCTCGCTGAC 95

RESULT 11
 CDS4UTM/c 970 bp DNA GSS 24-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone
 042K19 of library A from Tetraodon nigroviridis, genomic survey
 sequence.
 AL307831 GI:8213194
 VERSION GSS; genome survey sequence.
 KEYWORDS Tetraodon nigroviridis.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 970)
 Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 970)
 Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brothier,P., Queller,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 970)
 Genoscope.
 JOURNAL Direct Submission
 REFERENCE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 AUTHORS This sequence is a single read and was generated as part of a large
 COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 Location/Qualifiers
 1..970
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="042K19"
 /clone_lib="A"
 /note="Genoscope sequence ID : C0AA042AF10A3-end : T3"

BASE COUNT 262 a 221 c 245 g 213 t 29 others
 ORIGIN

Query Match 12.0%; Score 163.2; DB 13; Length 970;
 Best Local Similarity 56.1%; Pired. No. 1.9e-31;
 Matches 291; Conservative 0; Mismatches 228; Indels 0; Gaps 0;
 QY 204 gttatgaactgttcggtatctatgaacggagccttgaattggaagaagctc 263

Db 519 GTACACGTTAATTTGTCGCCGATATACAAATGACCCGATGAGGTGGAAATCTTT 460
 QY 264 ggaataaagaagaaggacatcatgaacttggaagatgatgttctggaatgacag 323
 Db 459 GNTCATCCGAGAGAAACCTCTGCCCGGAGTCCGACGAAACCTGACCAACTGACTN 400
 QY 324 tgatgtgacattatcagaccttaagaggttatgtctcaaaagctgtctcaagagaga 383
 Db 399 TGACTCAGAGAGCTTTGTGAGGCCACGATTCACGACAGAGAGTGCNTTCCGAATGCA 340
 QY 384 gaaagcttcccaatagcctatcttggtygttccaaaagatgcaatgatgttgaag 443
 Db 339 GCGAGANTTCCCTCGGCTACTCATATGTTGTGACANATATGCGTGTGATGTGAGAG 280
 QY 444 gttatcatgtcatataaacacagacacatattacatccatctatgacgtgaagc 503
 Db 279 GCTCCTCAGGCGCGGTACTCCCCCAATACATCTACTGCTCCACTAGACCTAAGTC 220
 QY 504 accgtatcctcaaaagtgtgcaataaacaattgaagtgcttctccaatatttcat 563
 Db 219 CCCCTACAGATTCAATTCGCGCATAGAGGCGCTGCTGCTGTCGCCCACTCTCAT 160
 QY 564 tgcctcaaatgagaagcgttggaatgacacatctccagacccagcgtgatgtaa 623
 Db 159 CGGCTCAAGCGGAGGTGTTCAATACGGGGCTTCAGCCGCTGAAAACCGACCTCA 100
 QY 624 ttgcttgcgaacctctgaagcttctcaatccagtggaatatgttatcaacttgtgtg 683
 Db 99 CTGCGCTGCGNCTTCTGAGGTCAAGGTCAAGTGAAGTACGATCAACACCTCTGGG 40
 QY 684 gcaagatttccctggaagctcaatttgaattgtgtgtc 722
 Db 39 CAGAGATTTCCTCCCTCAGGTCCCAACATCGNNTGNTNC 1

RESULT 12
 CDS10W2/c 756 bp DNA GSS 12-MAY-2000
 LOCUS Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
 197R05 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 AL168380 GI:7806437
 VERSION GSS; genome survey sequence.
 KEYWORDS Tetraodon nigroviridis.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 756)
 Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 756)
 Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brothier,P., Queller,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 756)
 Genoscope.
 JOURNAL Direct Submission
 REFERENCE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 AUTHORS This sequence is a single read and was generated as part of a large
 COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES

Location/Qualifiers
1. /756
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="197505"
/clone.lib="g"
/note="Genoscope sequence ID : C0AG197CC03Sp1-end :
puc-ori"

BASE COUNT 169 a 189 c 224 g 166 t 8 others

Query Match 11.6%; Score 157.6; DB 13; Length 756;
Best Local Similarity 61.2%; Pred. No. 5e-30;

Matches 267; Conservative 2; Mismatches 166; Indels 1; Gaps 1;

QY 848 taaggaacaacatctccaaggaagcaccaccataacatcagatattgttggcagtg 907
DB 754 TCAAAACGAGGTGGCCAAAGCCCGCCCGCCGACATCCAGKCTTCATCGCGACGC 695
QY 908 ctatttggtttaagcaagcatttgaataatatttcaacaactcattcgtcaag 967
DB 694 CTTATTGCTGTGTGCGGGGACTTTTGGCACACGTGAACAGAGGAGCTGGCCAGG 635
QY 968 acttttgcctgtctaaagacacatactctcgtatgagacatttgggtacttga 1027
DB 634 ACTTCCTGGCTGTGTCRGTGACACACTGCGGACGAGCACTTCTGGGCCACCTGG 575
QY 1028 ttgggttccaggaataactctggagagatttccagatcagccagagatgttcatctgc 1087
DB 574 TCAGGGTCCCGGGGGTCCCGCCCAATCCCGCTCCCAAGCCGATGACCGATGTA 515
QY 1088 agagtaagactcgccttgcagagtggaattactatgaaggcttttccatccagttgta 1147
DB 514 GAAGTAAGACGGCGGTGTGAATGGAATCTGTGAGAGGGAGAGCTGACCCGCCCTGCA 455
QY 1148 ctgtgtccacattcgaagctgtgtatttgaagctgcagaaattgaagtggtatca 1207
DB 454 CGGGACACACCTCGCGACCGGTGKCACTACGGGGCGCGGACTT-CGCTGGCTCTCG 396
QY 1208 aagatggacattgtttgctaataattgatttctaagtygagcctaattgaataa 1267
DB 395 GCTTGGAACACTGTTCCGCCAACAAGTTGAACCCCAAGTGAGCCCGGCTGTGATCAAGT 336
QY 1268 gcttggcagaagaact 1283
DB 335 GTTGGAGAGAGAGCT 320

RESULT 13
BE005007/c 526 bp mRNA EST 05-JUN-2000
LOCUS BE005007
DEFINITION MR0-BN0115-020300-001-803 BN0115 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE005007
VERSION BE005007.1 GI:8265240
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 526)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.R.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel.: +55-11-7704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612-MR0-BN0115-020
300-001-a03&rs=2000-03-02&t=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 526.

FEATURES

Location/Qualifiers
1. /526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="BN0115"
/dev.stage="Adult"
/note="Organ: breast, normal; Vector: puc18; Site:1: Smal;
Site:2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (O.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 126 a 125 c 136 g 139 t

Query Match 10.9%; Score 148.8; DB 10; Length 526;
Best Local Similarity 60.3%; Pred. No. 8.7e-28;
Matches 246; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 345 tctaagaggtatgctcaaaagcttctcaagagagagaaagcttccaatgctta 404
DB 497 TCGAAGCCACTATGTAAAGAAACACTCTGTGAAGAGGCTGGTTCCTTTAGCTTA 438
QY 405 ttcttgggtgtccacaagaatgcaattatgttgaaggtatccatgctatacaaa 464
DB 437 CACAGTACCATTCACAAAGACTTGGCACCTTTTGAAGCGGCTTTCAGGCGCATTTATAT 378
QY 465 ccagcaacaattactgcatcattatgatcgttaaggaacctgataacttcaagttgc 524
DB 377 GCCCCCAATAGTCTACTGTGTCACCTGTGACAGAGGCGACGATGCTTTAAAGTGC 318
QY 525 catgaacaattgctcgaagtgcttcccaatttccattgcttccaaattgaagctgt 584
DB 317 AGTAAACACTTACTACGCTGCTTCCCAATGCTTTTGTGCTTCCAAAGAGAGTGCST 258
QY 585 ggaatagccacattccagactccagctgatttaattgcttgcgaccttctga 644
DB 257 TGTCTATGGGGGATCTCCAGGCTCCAGCTGACCTGAACTGCTTGAAGACTTGTGCT 198
QY 645 gtcttcaatccagtggaatatttcaacctgtgtgggcaagatttccctgaagtc 704
DB 197 CTCTGAAGTTCCTGGAGATGATGATCAACACCCCGGCGGCAAGACTTTCCTGAAAC 138
QY 705 aaatttgaattgtgtcagagtgtgaaaaaaactcaatgagcaaatat 752
DB 137 CACAGGGAATAGTTCAGTATCTGAAGGAGATTAAAGGAAAAATAT 90

RESULT 14
BE292814 612 bp mRNA EST 13-JUL-2000
LOCUS BE292814
DEFINITION 6011051179F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:298090 5',
mRNA sequence.
ACCESSION BE292814
VERSION BE292814.1 GI:9175472
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 612)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
AUTHORS
NIH-MGC <http://mgc.nci.nih.gov/>
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at: image.lnl.gov
Plate: L10M76 row: m column: 19
High quality sequence stop: 606.
Location/Qualifiers

FEATURES
source

1. 612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH-MGC-15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site: 1: XhoI; Site 2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCGAGAG(c). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 181 a 137 c 140 g 154 t
ORIGIN

Query Match 10.3%; Score 148.8; DB 10; Length 612;
Best Local Similarity 54.0%; Pred. No. 8,9e-28;
Matches 340; Conservative 0; Mismatches 272; Indels 18; Gaps 1;

QY 395 caataagctatcttctgtgtcccaagaatgcaattatggttgaagagctatccatg 454
Db 1 CTATTGATACCTCTATGTGATTCATGAGAGATTGAAACCTTGAAAGGCTCTCGAG 60
QY 455 ctatacaaccagcacaatatctatgcatcatcattatgctgaagcaccgataccct 514
Db 61 CTGTGATGGCCCTCAGACATATACGTGTCCATGTGAGTAAGTCCCGCAGAACTT 120
QY 515 tcaagtgcaatgaacaattagctgaagtgcttcccaatttccattgtcttccaat 574
Db 121 TCAAGAGGCGGCTCAAGCAATATTCTTCTCCCAATGCTTCATAGCCAGTAAGC 180
QY 575 tagagagctgtgaatagccacattccagactccagagctgttgaattgtctcg 634
Db 181 TGTGTGCGGTGCTTTATGCTCTCGTCCAGGCTGCAAGTGCCTCAACTGATGGAAG 240
QY 635 acccttgaaagcttcaatccagtggaataatgtatcaactgtgtgaggaagatttc 694
Db 241 ACTTGCTCCAGAGCTCAGTCCCTGGAATATCTTCGATACATGTGGAGCGACTTTC 300
QY 695 cccctgaagtcacaatttgatgtgtgtcagagttgaaaaaacatcgaatgagcacaatgt 754
Db 301 CTATTAAGACCAATGCCAGATGCTCAGGCTCTCAAGATGTTGAATGGAGGAATAGCA 360
QY 755 tgaagagctgtgaaccccccaacagtaaatgtgaagatcacttccatcgaactta 814
Db 361 TGGAGTCAAGAGTACCTCTCAAGCAAAACCGCTGGAATATCACTTTGAGGTAG 420
QY 815 gaagtgagcttatgatatgtgaagctaccataagaagaacatcctcaagaagcag 874
Db 421 TGAAG-----ACACATTACACTTACCAACCAAGAAAGAGATCTTC 462
QY 875 ccccccaataatcagatatgtgttgagctgcttattgttttaagtcagacattg 934

Db 463 CCCCTATATATTACTATGATTACAGGGAATGCTCATATGTGCTTCCGAGATTTCG 522
QY 935 ttaatatatttcaacaacatccatcgttcaagacttttgcctgtctgaagacat 994
Db 523 TCCAACTGTGTTGAAGAACCCCTAAATCCCAACAACCTGATGAAAGGAGAACTT 582
QY 995 actcctcgtatgacacattgttgact 1024
Db 583 ATAGCCCAATGATACACCTCTGGCCACT 612

RESULT 15
BG185546 873 bp mRNA EST 21-APR-2001
LOCUS RST4495 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG185546
ACCESSION BG185546.1 GI:13707233
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
REFERENCE
AUTHORS
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashinsky,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Velloso,N., Hees,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.
and Ducar,M.
TITLE
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
JOURNAL
COMMENT
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scott@atersys.com
High quality sequence stop: 454.
Location/Qualifiers

1. 873
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 229 a 214 c 191 g 239 t
ORIGIN

Query Match 10.4%; Score 142; DB 11; Length 873;
Best Local Similarity 62.3%; Pred. No. 5,3e-26;
Matches 223; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 345 tctaaagagttatgtctcaaaagctgtctcaagaaggaagaagcttcccaatgcta 404
Db 374 TCGAAGCCACTGTTGAACGAAACACTCTTGAAAGAAAGCTGGTCCCTTATGCTTA 315
QY 405 tctctgtgtgtccacaagaatgcaattatgttgaaggttaccatccatatacaaa 464
Db 314 CACAGTGAACCAATCCACAAAGACTTGGCACTTTGAGAGGCTTCAAGGGGATTTAT 255
QY 465 ccagacacattattctgcatcattatgcatcgttaagcaccctgataactcacaagtgc 524
Db 254 GCCCAAAATGCTACTGTGTCACCTGATCAGAGGCGAGGATGCTTTAAAGGTGC 195
QY 525 catgacaattagctaaagtgcttcccaatatttccattcttcccaattagagctgt 584

Fri May 3 10:57:54 2002

us-09-645-192-1.rst

Page 12

Db 194 AGTAAACAGTACTCAGCGCTCTCCCAATAGCTTTCTGCGCTCCACAGAGAGAGTGGCT 135

QY 585 ggaatattggcccaactttcttcagacacacacggcgatattaaattgccttgcgagccttcgaa 644

Db 134 TGTATTATGGGGGATCTTCACAGCTCCAGGCTCAGACTGAACCTGGTGAACACCTTTGGGC 75

QY 645 gtcttcaatccagaggaatatgtatcaacttggggcagaattttccctcgaag 702

Db 74 CTCGTAAAGTTCCTCGAAGATATGTCATCAACACCTGGCGGCAACCTTTCCCTGTAG 17

Search completed: May 2, 2002, 09:55:22
Job time: 5812 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 10:36:36 ; Search time 1980.83 seconds
(without alignments)
11343.305 Million cell updates/sec

Title: US-09-645-192-1

Perfect score: 1362
Sequence: 1 atgaagatattcaatgtta.....atctcaactaccacatcatga 1362

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 10

Total number of hits satisfying chosen parameters: 1358776

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1362	100.0	1362	6	AX087935 Sequence
2	1362	100.0	3435	2	AF133035
3	1260	92.5	184590	2	AC093259 Homo sapi
4	25	1.8	31	6	AX087939 Sequence
5	21	1.5	21	6	AX087937 Sequence
6	21	1.5	696	8	CNS01AY6
7	21	1.5	9643	2	AC014188
8	21	1.5	24650	2	AC023731
9	21	1.5	84011	2	AP002094
10	21	1.5	133008	2	AC011375
11	21	1.5	146190	2	AC073826
12	21	1.5	155359	2	AL358777 Human DNA
13	21	1.5	161676	2	AC023461
14	21	1.5	178098	9	HS168L15
15	21	1.5	188353	9	AC009480
16	21	1.5	195227	2	AL592214
17	21	1.5	196788	2	AC021626
18	21	1.5	301051	3	AE003487
19	20	1.5	32	6	AX087938
20	20	1.5	173	9	HUMPSABD
21	20	1.5	185	9	HUMPRBS11
22	20	1.5	478	9	HUMSPS08
23	20	1.5	732	10	MUSCYCP32
24	20	1.5	1086	14	FLBRABD
25	20	1.5	2165	1	EU059304
26	20	1.5	2705	10	AF327184S1
27	20	1.5	61274	9	AL391815
28	20	1.5	70339	9	AL513123
29	20	1.5	82101	9	AC005694
30	20	1.5	89862	9	AC007164
31	20	1.5	110443	9	AL354692
32	20	1.5	120891	2	AC011285
33	20	1.5	126138	9	HSAR18000
34	20	1.5	156374	9	AC007559
35	20	1.5	162323	2	AL596283
36	20	1.5	163189	2	AC073556
37	20	1.5	170294	2	AC016845
38	20	1.5	173899	9	AC010150
39	20	1.5	174596	2	AC023812
40	20	1.5	174725	9	AC007486
41	20	1.5	175092	2	AC013261
42	20	1.5	181994	2	AC090822
43	20	1.5	183807	2	AC026544
44	20	1.5	183147	2	AC080188
45	20	1.5	185421	2	AC079856
46	20	1.5	198105	2	AC026160
47	20	1.5	235281	9	AC004673
48	20	1.5	318488	9	AC005529
49	19	1.4	496	11	G55641
50	19	1.4	1087	14	AF217223
51	19	1.4	1196	8	HS1RPS05
52	19	1.4	1752	8	D89126
53	19	1.4	2835	8	SCSCY13GN
54	19	1.4	2857	1	AF278705
55	19	1.4	3065	8	HS292867
56	19	1.4	3528	8	SCYGL083W
57	19	1.4	3535	8	YSPCTL
58	19	1.4	11649	1	AE001880
59	19	1.4	33450	2	CEH27104
60	19	1.4	33869	3	CEH38C2
61	19	1.4	36258	3	CEH29F11
62	19	1.4	37194	8	HSV210E9
63	19	1.4	37228	8	SPCC757
64	19	1.4	40600	3	CEH20D1
65	19	1.4	43349	9	HSAC000367
66	19	1.4	47079	2	AC083776
67	19	1.4	51831	9	AL138999
68	19	1.4	52290	8	AC025314
69	19	1.4	66136	8	AB024031
70	19	1.4	70169	8	AF296833 Arabidops

71	19	1.4	71302	2	AC087760	AC087760 Homo sapi	C 144	19	1.4	188613	2	AC012133	AC012133 Homo sapi
72	19	1.4	78260	2	AC015439	AC015439 Drosophila	C 145	19	1.4	188643	2	AC092422	AC092422 Homo sapi
73	19	1.4	83656	2	AC007401	AC007401 Homo sapi	C 146	19	1.4	189032	2	AC025564	AC025564 Homo sapi
74	19	1.4	85832	9	HS344P17	Z95326 Human DNA s	C 147	19	1.4	189602	2	AC023237	AC023237 Homo sapi
75	19	1.4	86516	9	AC009493	AC009493 Homo sapi	C 148	19	1.4	190166	33	AC026285	AC026285 Homo sapi
76	19	1.4	89490	8	TM15	AF296832 Arabidops	C 149	19	1.4	190548	2	AC027677	AC027677 Homo sapi
77	19	1.4	89795	8	NC3H10	AL513442 Neurospor	C 150	19	1.4	190773	9	AC018513	AC018513 Homo sapi
78	19	1.4	101270	2	HS483K16	AL034374 Human DNA	C 151	19	1.4	191570	2	AL591598	AL591598 Mus muscu
79	19	1.4	101555	2	AF188026	AF188026 Homo sapi	C 152	19	1.4	192169	2	AC020849	AC020849 Mus muscu
80	19	1.4	104810	9	AC003012	AC003012 Human PAC	C 153	19	1.4	192422	2	AC021071	AC021071 Homo sapi
81	19	1.4	124823	9	AC007064	AC007064 Homo sapi	C 154	19	1.4	194026	2	AC084313	AC084313 Homo sapi
82	19	1.4	125856	9	HS1187M17	AL121891 Human DNA	C 155	19	1.4	194734	2	AC011386	AC011386 Homo sapi
83	19	1.4	126736	2	AC044788	AC044788 Homo sapi	C 156	19	1.4	196375	2	AC073107	AC073107 Homo sapi
84	19	1.4	129517	2	AC016400	AC016400 Homo sapi	C 157	19	1.4	198146	2	AC023499	AC023499 Homo sapi
85	19	1.4	130363	2	AC006346	AC006346 Homo sapi	C 158	19	1.4	198748	2	AC016719	AC016719 Homo sapi
86	19	1.4	132183	2	AC011376	AC011376 Homo sapi	C 159	19	1.4	199255	2	AC096720	AC096720 Homo sapi
87	19	1.4	136158	9	HS1049G11	AL078634 Human DNA	C 160	19	1.4	199702	2	AC087178	AC087178 Homo sapi
88	19	1.4	139079	2	AC022195	AC022195 Homo sapi	C 161	19	1.4	200149	2	AC006548	AC006548 Homo sapi
89	19	1.4	140864	2	AC074355	AC074355 Oryza sat	C 162	19	1.4	203449	2	AC073210	AC073210 Homo sapi
90	19	1.4	143094	2	AL591114	AL591114 Homo sapi	C 163	19	1.4	212549	9	AC009492	AC009492 Homo sapi
91	19	1.4	143549	2	AP003910	AP003910 Oryza sat	C 164	19	1.4	214631	2	AC060805	AC060805 Homo sapi
92	19	1.4	145520	2	AC016036	AC016036 Homo sapi	C 165	19	1.4	218470	1	BSUB0013	BSUB0013 Mus muscu
93	19	1.4	147052	2	AC090687	AC090687 Homo sapi	C 166	19	1.4	220735	2	AC087889	AC087889 Mus muscu
94	19	1.4	147619	9	AL512358	AL512358 Human DNA	C 167	19	1.4	224076	2	NC084015	NC084015 Mus muscu
95	19	1.4	149903	2	AC090510	AC090510 Homo sapi	C 168	19	1.4	282700	1	BACJH642	BACJH642 Homo sapi
96	19	1.4	150650	2	AC064845	AC064845 Homo sapi	C 169	19	1.4	317511	3	CUI1168X3	AL133076 Campyloba
97	19	1.4	154028	2	HSAC002087	AC002087 Human BAC	C 170	19	1.4	342068	6	AE003440	AE003440 Drosophila
98	19	1.4	154561	2	AC087344	AC087344 Homo sapi	C 171	19	1.3	342068	6	AX087940	AX087940 Drosophila
99	19	1.4	154811	2	AC024699	AC024699 Homo sapi	C 172	19	1.3	250	11	AY021388	AY021388 Oryza sat
100	19	1.4	154953	2	AL591023	AL591023 Homo sapi	C 173	19	1.3	250	11	G15516	G15516 human STS S
101	19	1.4	155021	9	AC007759	AC007759 Homo sapi	C 174	19	1.3	297	5	HEFICRCH1	M17166 Shark IGH C
102	19	1.4	155026	2	AL442128	AL442128 Human DNA	C 175	19	1.3	297	5	HEFICRCH1	M17167 Shark IGH C
103	19	1.4	156301	2	AC083900	AC083900 Homo sapi	C 176	19	1.3	592	6	A96052	A96052 Sequence 85
104	19	1.4	156596	2	AL355479	AL355479 Homo sapi	C 177	19	1.3	599	9	AF098981	AF098981 Oxalis sp
105	19	1.4	157340	2	AC015652	AC015652 Homo sapi	C 178	19	1.3	740	11	G64409	G64409 B616B4/Sp6
106	19	1.4	158391	2	AC008023	AC008023 Homo sapi	C 179	19	1.3	780	6	RCU14374	RCU14374 Rana catesbe
107	19	1.4	159517	2	AC018511	AC018511 Homo sapi	C 180	19	1.3	807	6	A96054	A96054 Sequence 87
108	19	1.4	159743	2	AC011357	AC011357 Homo sapi	C 181	19	1.3	807	6	A96056	A96056 Sequence 89
109	19	1.4	160274	2	AC091953	AC091953 Homo sapi	C 182	19	1.3	807	6	A96056	A96056 Sequence 93
110	19	1.4	160422	2	AC011087	AC011087 Homo sapi	C 183	19	1.3	903	6	A96056	A96056 Sequence 93
111	19	1.4	162505	2	AC024258	AC024258 Homo sapi	C 184	19	1.3	903	6	A96056	A96056 Sequence 93
112	19	1.4	162593	2	AC022576	AC022576 Homo sapi	C 185	19	1.3	919	11	CNS06K78	AL166485 Glycine m
113	19	1.4	162875	2	AC018509	AC018509 Homo sapi	C 186	19	1.3	995	11	CNS06K78	AL166485 Glycine m
114	19	1.4	163404	2	AL355308	AL355308 Homo sapi	C 187	19	1.3	1047	5	AE111218	AE111218 Cyrtineil
115	19	1.4	164405	2	AL513488	AL513488 Homo sapi	C 188	19	1.3	1085	14	TM9REPPPO	M90495 Tomato molt
116	19	1.4	166919	2	AL591675	AL591675 Mus muscu	C 189	19	1.3	1251	8	HEFICRCH1	AF3782 Heterodontu
117	19	1.4	166971	9	AL157812	AL157812 Human DNA	C 190	19	1.3	1457	8	AF349521	AF349521 Arabidops
118	19	1.4	168025	9	AC024364	AC024364 Homo sapi	C 191	19	1.3	1755	8	AF327420	AF327420 Arabidops
119	19	1.4	168278	2	AC037492	AC037492 Homo sapi	C 192	19	1.3	2475	9	EGTUB2	X1326 Erysiphe gr
120	19	1.4	169268	2	AC027674	AC027674 Homo sapi	C 193	19	1.3	2745	10	BSA252060	BSA252060 Homo sapi
121	19	1.4	171490	2	AC068209	AC068209 Homo sapi	C 194	19	1.3	2858	10	BC006700	BC006700 Mus muscu
122	19	1.4	171690	2	AC025925	AC025925 Homo sapi	C 195	19	1.3	3069	10	SUSVEBIV	BC006700 Mus muscu
123	19	1.4	171764	2	AC022803	AC022803 Homo sapi	C 196	19	1.3	3543	6	SAPKNH	LI2536 Sea urchin
124	19	1.4	172749	2	AC068042	AC068042 Homo sapi	C 197	19	1.3	3664	6	AY056399	AY056399 Sequence
125	19	1.4	173268	9	AL138760	AL138760 Human DNA	C 198	19	1.3	3958	6	YC058M6A	YC058M6A Bombyx mo
126	19	1.4	174437	2	AC007768	AC007768 Homo sapi	C 199	19	1.3	4116	3	AB060286	AB060286 Bombyx mo
127	19	1.4	175020	9	AC022317	AC022317 Homo sapi	C 200	19	1.3	4240	3	BM0X0	BM0X0 Bombyx mo
128	19	1.4	175265	9	AC063964	AC063964 Homo sapi	C 201	19	1.3	4866	8	YC058M6A	YC058M6A Bombyx mo
129	19	1.4	176630	2	AC010684	AC010684 Homo sapi	C 202	19	1.3	5313	8	MIZMRPBG	MIZMRPBG Bombyx mo
130	19	1.4	177407	2	AC009880	AC009880 Homo sapi	C 203	19	1.3	5332	6	HSW802808	HSW802808 Homo sapi
131	19	1.4	178597	2	AC063962	AC063962 Homo sapi	C 204	19	1.3	5355	6	E30078	E30078 Novel prote
132	19	1.4	180568	2	AC006482	AC006482 Homo sapi	C 205	19	1.3	5355	6	E30080	E30080 Novel prote
133	19	1.4	180707	2	AC024460	AC024460 Homo sapi	C 206	19	1.3	5374	8	SCYBRI12C	SCYBRI12C Homo sapi
134	19	1.4	181902	2	AC020788	AC020788 Homo sapi	C 207	19	1.3	7360	8	HEFICRCH1	HEFICRCH1 Homo sapi
135	19	1.4	183938	2	AC023876	AC023876 Homo sapi	C 208	19	1.3	7500	8	RICMTRP	RICMTRP Homo sapi
136	19	1.4	184804	2	AC036174	AC036174 Homo sapi	C 209	19	1.3	8237	8	AWU10505	AWU10505 Homo sapi
137	19	1.4	185018	2	AL439602	AL439602 Homo sapi	C 210	19	1.3	8949	8	SCMDLX5	SCMDLX5 Homo sapi
138	19	1.4	185838	9	AC036191	AC036191 Homo sapi	C 211	19	1.3	9971	1	U67577	U67577 Methanococ
139	19	1.4	186298	2	AC007860	AC007860 Homo sapi	C 212	19	1.3	11730	1	AE002480	AE002480 Neisseria
140	19	1.4	186360	2	AL512429	AL512429 Homo sapi	C 213	19	1.3	16781	9	AL536422	AL536422 Homo sapi
141	19	1.4	186526	2	AC026130	AC026130 Homo sapi	C 214	19	1.3	17385	8	OSA277468	OSA277468 Oryza sat
142	19	1.4	187020	2	AC026614	AC026614 Homo sapi	C 215	19	1.3	17769	9	AL391814	AL391814 Homo sapi
143	19	1.4	187894	9	AC007320	AC007320 Homo sapi	C 216	19	1.3	21115	3	AB005911	AB005911 Bombyx mo

C 217	18	1.3	27333	3	CELT28B4	AF026206 Caenorhab	C 290	18	1.3	122537	2	AP003858	AP003858 Oryza sat
C 218	18	1.3	28585	3	U41031	U41031 Caenorhabd1	C 291	18	1.3	123778	2	AC069294	AC069294 Homo sapi
C 219	18	1.3	29605	3	AB020747	AB020747 Arabidops	C 292	18	1.3	124539	2	AP003166	AP003166 Homo sapi
C 220	18	1.3	31537	2	AC091528.4	Continuation (5 of	C 293	18	1.3	125397	2	AC008550	AC008550 Homo sapi
C 221	18	1.3	33649	3	CBRG44J05	AC084618 Caenorhab	C 294	18	1.3	125595	2	AL139149	AL139149 Homo sapi
C 222	18	1.3	35241	3	CELY39A3B	AC006748 Caenorhab	C 295	18	1.3	126038	8	AP000367	AP000367 Homo sapi
C 223	18	1.3	37526	3	AL139825	AL139825 Human DNA	C 296	18	1.3	126039	8	AP0003187	AP0003187 Oryza sat
C 224	18	1.3	39339	3	CEK08H10	283113 Caenorhabd1	C 297	18	1.3	126839	8	HS19408	HS19408 Homo sapi
C 225	18	1.3	39439	3	AF067219	AF067219 Caenorhab	C 298	18	1.3	127066	2	AC084265	AC084265 Homo sapi
C 226	18	1.3	42503	3	HSU115G11	271187 Human DNA s	C 299	18	1.3	129090	2	AP003931	AP003931 Oryza sat
C 227	18	1.3	43138	3	AL590100	AL590100 Human DNA	C 300	18	1.3	129263	9	AL360089	AL360089 Human DNA
C 228	18	1.3	43526	3	CELC10H11	U88311 Caenorhabd1	C 301	18	1.3	129355	9	AC004853	AC004853 Homo sapi
C 229	18	1.3	46239	1	AC027138	AC027138 Staphyloc	C 302	18	1.3	129422	2	AL589677	AL589677 Human DNA
C 230	18	1.3	46848	2	AC013885	AC013885 Drosophi1	C 303	18	1.3	131273	2	AC016441	AC016441 Homo sapi
C 231	18	1.3	51730	2	AC019578	AC019578 Drosophi1	C 304	18	1.3	131458	9	AC002386	AC002386 Human BAC
C 232	18	1.3	53546	2	AC084346	AC084346 Homo sapi	C 305	18	1.3	133787	9	AC015540	AC015540 Homo sapi
C 233	18	1.3	61226	2	AL139392	AL139392 Human DNA	C 306	18	1.3	134464	10	AC084052	AC084052 Mus Muscu
C 234	18	1.3	63743	2	AC073401	AC073401 Homo sapi	C 307	18	1.3	136129	2	AC011050	AC011050 Homo sapi
C 235	18	1.3	66424	2	AL512509	AL512509 Human DNA	C 308	18	1.3	136672	2	AC092451	AC092451 Homo sapi
C 236	18	1.3	69748	8	SCRACIT	X78993 S.cerevisia	C 309	18	1.3	137032	2	AC007481	AC007481 Homo sapi
C 237	18	1.3	69787	2	AC003984	AC003984 Homo sapi	C 310	18	1.3	137393	2	AC087318	AC087318 Homo sapi
C 238	18	1.3	69961	2	AC008419	AC008419 Homo sapi	C 311	18	1.3	137526	2	AC090683	AC090683 Oryza sat
C 239	18	1.3	71000	8	SEBP8B7	AC0042684 S.pombe	C 312	18	1.3	137659	9	AL139232	AL139232 Human DNA
C 240	18	1.3	73371	2	AC004872	AC004872 Homo sapi	C 313	18	1.3	137676	9	AP001607	AP001607 Homo sapi
C 241	18	1.3	75073	2	CEY48A6B	AC087648 Homo sapi	C 314	18	1.3	137879	8	AP003233	AP003233 Oryza sat
C 242	18	1.3	75073	2	CEY48A6B	AL023844 Caenorhab	C 315	18	1.3	138430	2	AL592204	AL592204 Danio rer
C 243	18	1.3	75935	2	AC022818	AC022818 Homo sapi	C 316	18	1.3	138459	2	AP003936	AP003936 Oryza sat
C 244	18	1.3	76825	2	AL159165	AL159165 Homo sapi	C 317	18	1.3	138633	2	AL356360	AL356360 Homo sapi
C 245	18	1.3	79561	2	AL135799	AL135799 Homo sapi	C 318	18	1.3	139152	6	AP002525	AP002525 Oryza sat
C 246	18	1.3	79874	2	AC027252	AC027252 Homo sapi	C 319	18	1.3	140229	8	AP003143	AP003143 Oryza sat
C 247	18	1.3	81369	6	AX034587	AX034587 Sequence	C 320	18	1.3	140899	2	AC083769	AC083769 Homo sapi
C 248	18	1.3	82456	2	HS153G14	AL031118 Human DNA	C 321	18	1.3	140952	8	AP003046	AP003046 Oryza sat
C 249	18	1.3	84096	2	AC074170	AC074170 Mus muscu	C 322	18	1.3	141432	2	AC015699	AC015699 Homo sapi
C 250	18	1.3	84570	2	AL359552	AL359552 Human DNA	C 323	18	1.3	141475	2	AC002980	AC002980 Homo sapi
C 251	18	1.3	86828	2	AC020569	AC020569 Homo sapi	C 324	18	1.3	141561	2	AC008979	AC008979 Homo sapi
C 252	18	1.3	87394	8	AC005882	AC005882 Arabidops	C 325	18	1.3	143291	2	AC026414	AC026414 Homo sapi
C 253	18	1.3	90220	2	HSJ734P14	AL043650 Human DNA	C 326	18	1.3	143710	2	AP002899	AP002899 Oryza sat
C 254	18	1.3	90935	2	AF003917	AF003917 Oryza sat	C 327	18	1.3	144314	8	AP000835	AP000835 Homo sapi
C 255	18	1.3	91893	8	AC005168	AC005168 Arabidops	C 328	18	1.3	144319	9	AC007968	AC007968 Homo sapi
C 256	18	1.3	92388	9	AC008802	AC008802 Homo sapi	C 329	18	1.3	144324	2	AC022662	AC022662 Homo sapi
C 257	18	1.3	92817	2	HSJ393D12	AL132776 Human DNA	C 330	18	1.3	144440	2	AC037436	AC037436 Homo sapi
C 258	18	1.3	93153	2	AF257499	AF257499 Homo sapi	C 331	18	1.3	144524	2	AC013457	AC013457 Homo sapi
C 259	18	1.3	94779	2	AP000664	AP000664 Homo sapi	C 332	18	1.3	146118	6	AP002521	AP002521 Oryza sat
C 260	18	1.3	95039	2	AC010068	AC010068 Drosophi1	C 333	18	1.3	146335	8	AP002521	AP002521 Oryza sat
C 261	18	1.3	95546	9	AC022816	AC022816 Homo sapi	C 334	18	1.3	146360	9	HS1156M12	HS1156M12 Homo sapi
C 262	18	1.3	96090	2	AP003908	AP003908 Oryza sat	C 335	18	1.3	146937	2	AL592423	AL592423 Homo sapi
C 263	18	1.3	96180	9	AC005868	AC005868 Homo sapi	C 336	18	1.3	147035	2	AP003418	AP003418 Oryza sat
C 264	18	1.3	97656	9	AP000629	AP000629 Homo sapi	C 337	18	1.3	147038	2	AL133326	AL133326 Human DNA
C 265	18	1.3	99908	2	AC006464	AC006464 Homo sapi	C 338	18	1.3	147117	8	AC069300	AC069300 Oryza sat
C 266	18	1.3	100906	8	ATF24G16	AL138647 Arabidops	C 339	18	1.3	147334	2	AP003349	AP003349 Oryza sat
C 267	18	1.3	102378	2	AP003736	AP003736 Oryza sat	C 340	18	1.3	147372	2	AC018538	AC018538 Homo sapi
C 268	18	1.3	105686	2	NC012072	NC012072 Homo sapi	C 341	18	1.3	147452	2	AL358781	AL358781 Homo sapi
C 269	18	1.3	105769	2	AL592429	AL592429 Danio rer	C 342	18	1.3	147518	33	AC027518	AC027518 Homo sapi
C 270	18	1.3	107101	2	AC008520	AC008520 Homo sapi	C 343	18	1.3	147712	9	AC011299	AC011299 Homo sapi
C 271	18	1.3	108399	2	AL356736	AL356736 Homo sapi	C 344	18	1.3	147741	2	AC092629	AC092629 Homo sapi
C 272	18	1.3	108569	2	AC008837	AC008837 Homo sapi	C 345	18	1.3	147865	2	AL591402	AL591402 Homo sapi
C 273	18	1.3	108634	2	AC020813	AC020813 Mus muscu	C 346	18	1.3	148018	2	AC068861	AC068861 Homo sapi
C 274	18	1.3	109074	2	AP000706	AP000706 Homo sapi	C 347	18	1.3	148454	9	AC007023	AC007023 Homo sapi
C 275	18	1.3	110000	2	AC090750.3	Continuation (4 of	C 348	18	1.3	148762	8	AP002843	AP002843 Oryza sat
C 276	18	1.3	110000	2	AC091528.3	Continuation (4 of	C 349	18	1.3	148920	2	AP003343	AP003343 Oryza sat
C 277	18	1.3	110000	2	AL162501.1	Continuation (2 of	C 350	18	1.3	148931	2	AC034112	AC034112 Homo sapi
C 278	18	1.3	110000	2	AL359978.0	Continuation (2 of	C 351	18	1.3	149618	9	AP000556	AP000556 Homo sapi
C 279	18	1.3	110000	2	AL359978.1	Continuation (2 of	C 352	18	1.3	149901	2	AC022243	AC022243 Homo sapi
C 280	18	1.3	110403	2	AC058813	AC058813 Homo sapi	C 353	18	1.3	150036	8	AP000557	AP000557 Homo sapi
C 281	18	1.3	110952	2	AL136369	AL136369 Human DNA	C 354	18	1.3	150150	8	AP002539	AP002539 Oryza sat
C 282	18	1.3	111182	9	AC073608	AC073608 Homo sapi	C 355	18	1.3	150379	8	AP003074	AP003074 Oryza sat
C 283	18	1.3	112366	9	AL591804	AL591804 Human DNA	C 356	18	1.3	150519	8	AP002993	AP002993 Lotus Jap
C 284	18	1.3	112884	9	AL136966	AL136966 Human DNA	C 357	18	1.3	150583	9	AC026884	AC026884 Homo sapi
C 285	18	1.3	112902	9	AC005230	AC005230 Homo sapi	C 358	18	1.3	150814	2	AP003541	AP003541 Oryza sat
C 286	18	1.3	113515	9	HSJ944F13	HSJ944F13 Human DNA	C 359	18	1.3	150849	2	AC020890	AC020890 Homo sapi
C 287	18	1.3	117757	9	HSJ100419	AL121912 Human DNA	C 360	18	1.3	151321	2	AC012276	AC012276 Homo sapi
C 288	18	1.3	118653	9	AP003738	AP003738 Oryza sat	C 361	18	1.3	151335	2	AC023302	AC023302 Homo sapi
C 289	18	1.3	119972	2	AP004029	AP004029 Oryza sat	C 362	18	1.3	151340	2	AC069132	AC069132 Homo sapi

C 363	18	1.3	151755	2	AC013299	AC013299 Homo sapi	C 436	18	1.3	165262	2	AP003181	AP003181 Homo sapi
C 364	18	1.3	152127	2	AC060759	AC060759 Homo sapi	C 437	18	1.3	165289	2	HS45P15	AL442638 Homo sapi
C 365	18	1.3	152582	2	AC027014	AC027014 Homo sapi	C 438	18	1.3	165388	2	AL590966	AL590966 Homo sapi
C 366	18	1.3	152935	2	AC012642	AC012642 Homo sapi	C 439	18	1.3	165860	9	AP000961	AP000961 Homo sapi
C 367	18	1.3	152959	2	AC011275	AC011275 Homo sapi	C 440	18	1.3	165915	2	AC092922	AC092922 Homo sapi
C 368	18	1.3	153215	2	HS24A17	AL035452 Homo sapi	C 441	18	1.3	166071	9	AC022336	AC022336 Homo sapi
C 369	18	1.3	153289	2	AC013553	AC013553 Homo sapi	C 442	18	1.3	166287	2	AC026083	AC026083 Homo sapi
C 370	18	1.3	153408	2	AC011995	AC011995 Homo sapi	C 443	18	1.3	166620	2	AC020630	AC020630 Homo sapi
C 371	18	1.3	153764	9	AL355350	AL355350 Homo sapi	C 444	18	1.3	166782	2	AC021516	AC021516 Homo sapi
C 372	18	1.3	154091	9	AP003819	AP003819 Homo sapi	C 445	18	1.3	167025	2	AC021380	AC021380 Homo sapi
C 373	18	1.3	154172	9	AC079386	AC079386 Homo sapi	C 446	18	1.3	167163	2	AC015705	AC015705 Homo sapi
C 374	18	1.3	154619	1	D90917	D90917 Synchocyst	C 447	18	1.3	167437	2	AC023068	AC023068 Homo sapi
C 375	18	1.3	155278	2	AC040161	AC040161 Homo sapi	C 448	18	1.3	167631	9	AL158829	AL158829 Homo sapi
C 376	18	1.3	155576	2	AC055712	AC055712 Homo sapi	C 449	18	1.3	167862	9	AC011966	AC011966 Homo sapi
C 377	18	1.3	155674	2	AC073140	AC073140 Homo sapi	C 450	18	1.3	168293	2	AL451137	AL451137 Homo sapi
C 378	18	1.3	155774	2	AL592202	AL592202 Homo sapi	C 451	18	1.3	168334	2	AC027208	AC027208 Homo sapi
C 379	18	1.3	155804	2	AC019241	AC019241 Homo sapi	C 452	18	1.3	168467	2	AP001143	AP001143 Homo sapi
C 380	18	1.3	156002	2	AC036110	AC036110 Homo sapi	C 453	18	1.3	168538	2	AC091722	AC091722 Homo sapi
C 381	18	1.3	156069	2	AC079256	AC079256 Homo sapi	C 454	18	1.3	168561	2	AC091989	AC091989 Homo sapi
C 382	18	1.3	156334	2	AC025755	AC025755 Homo sapi	C 455	18	1.3	168623	2	AC084013	AC084013 Homo sapi
C 383	18	1.3	156403	2	AC078823	AC078823 Homo sapi	C 456	18	1.3	168624	33	AC021874	AC021874 Homo sapi
C 384	18	1.3	156457	2	AC093014	AC093014 Homo sapi	C 457	18	1.3	168720	2	AC092987	AC092987 Homo sapi
C 385	18	1.3	156457	9	AC004383	AC004383 Homo sapi	C 458	18	1.3	168777	2	AL356773	AL356773 Homo sapi
C 386	18	1.3	156485	2	AL359701	AL359701 Homo sapi	C 459	18	1.3	168893	2	AC032018	AC032018 Homo sapi
C 387	18	1.3	156724	2	AC016306	AC016306 Homo sapi	C 460	18	1.3	168990	2	AC079324	AC079324 Homo sapi
C 388	18	1.3	156957	2	AP001977	AP001977 Homo sapi	C 461	18	1.3	169237	9	AC009516	AC009516 Homo sapi
C 389	18	1.3	157074	2	AC055708	AC055708 Homo sapi	C 462	18	1.3	169337	2	AL356462	AL356462 Homo sapi
C 390	18	1.3	157077	2	AC022504	AC022504 Homo sapi	C 463	18	1.3	169401	2	AC059056	AC059056 Homo sapi
C 391	18	1.3	157086	2	AP000552	AP000552 Homo sapi	C 464	18	1.3	169939	2	AC079765	AC079765 Homo sapi
C 392	18	1.3	157183	9	AC011731	AC011731 Homo sapi	C 465	18	1.3	170224	2	AL139341	AL139341 Homo sapi
C 393	18	1.3	157358	2	AC024906	AC024906 Homo sapi	C 466	18	1.3	170508	2	AC022450	AC022450 Homo sapi
C 394	18	1.3	157348	2	AC021972	AC021972 Homo sapi	C 467	18	1.3	170602	2	AC022536	AC022536 Homo sapi
C 395	18	1.3	158026	2	AL158213	AL158213 Homo sapi	C 468	18	1.3	170796	2	AC069416	AC069416 Homo sapi
C 396	18	1.3	158039	2	AC021840	AC021840 Homo sapi	C 469	18	1.3	170934	2	AL354684	AL354684 Homo sapi
C 397	18	1.3	158148	2	AC015463	AC015463 Homo sapi	C 470	18	1.3	170939	2	AC069536	AC069536 Homo sapi
C 398	18	1.3	158236	9	AC011631	AC011631 Homo sapi	C 471	18	1.3	170995	2	AC092040	AC092040 Homo sapi
C 399	18	1.3	158723	8	AP002869	AP002869 Homo sapi	C 472	18	1.3	171044	2	AC092287	AC092287 Homo sapi
C 400	18	1.3	159160	33	AC021023	AC021023 Homo sapi	C 473	18	1.3	171073	2	AC073380	AC073380 Homo sapi
C 401	18	1.3	159217	9	AC016950	AC016950 Homo sapi	C 474	18	1.3	171185	2	AL365507	AL365507 Homo sapi
C 402	18	1.3	159222	2	AC092419	AC092419 Homo sapi	C 475	18	1.3	171206	2	AC026573	AC026573 Homo sapi
C 403	18	1.3	159287	2	AC079119	AC079119 Homo sapi	C 476	18	1.3	171245	9	AC044791	AC044791 Homo sapi
C 404	18	1.3	159387	2	AC010162	AC010162 Homo sapi	C 477	18	1.3	171490	33	AC036166	AC036166 Homo sapi
C 405	18	1.3	159453	2	AL596267	AL596267 Homo sapi	C 478	18	1.3	171523	2	AC010206	AC010206 Homo sapi
C 406	18	1.3	159586	2	AF377946	AF377946 Homo sapi	C 479	18	1.3	171621	2	AC011961	AC011961 Homo sapi
C 407	18	1.3	159586	2	AF377946	AF377946 Homo sapi	C 480	18	1.3	171828	2	AC022247	AC022247 Homo sapi
C 408	18	1.3	159637	2	AC007990	AC007990 Homo sapi	C 481	18	1.3	172571	2	AC022748	AC022748 Homo sapi
C 409	18	1.3	159816	2	AC011075	AC011075 Homo sapi	C 482	18	1.3	172747	9	AL359835	AL359835 Homo sapi
C 410	18	1.3	159836	2	AC016701	AC016701 Homo sapi	C 483	18	1.3	172792	9	AC062031	AC062031 Homo sapi
C 411	18	1.3	160196	2	AC025005	AC025005 Homo sapi	C 484	18	1.3	173141	2	AC027508	AC027508 Homo sapi
C 412	18	1.3	160421	2	AL512584	AL512584 Homo sapi	C 485	18	1.3	173445	2	AL162380	AL162380 Homo sapi
C 413	18	1.3	160541	2	AC007944	AC007944 Homo sapi	C 486	18	1.3	173465	2	AC084374	AC084374 Homo sapi
C 414	18	1.3	160760	2	AC079772	AC079772 Homo sapi	C 487	18	1.3	173506	2	AC016323	AC016323 Homo sapi
C 415	18	1.3	160993	2	AC018407	AC018407 Homo sapi	C 488	18	1.3	173521	33	AC067850	AC067850 Homo sapi
C 416	18	1.3	161394	2	AC010759	AC010759 Homo sapi	C 489	18	1.3	173560	2	AL513262	AL513262 Homo sapi
C 417	18	1.3	161459	2	AC023594	AC023594 Homo sapi	C 490	18	1.3	173764	2	AC012464	AC012464 Homo sapi
C 418	18	1.3	161459	2	AC023594	AC023594 Homo sapi	C 491	18	1.3	173879	2	AC092600	AC092600 Homo sapi
C 419	18	1.3	161687	2	AC048351	AC048351 Homo sapi	C 492	18	1.3	173891	2	AC009545	AC009545 Homo sapi
C 420	18	1.3	161876	2	AC073634	AC073634 Homo sapi	C 493	18	1.3	173987	33	AC027165	AC027165 Homo sapi
C 421	18	1.3	162194	2	AC026592	AC026592 Homo sapi	C 494	18	1.3	174095	2	AL356788	AL356788 Homo sapi
C 422	18	1.3	162346	9	AC004551	AC004551 Homo sapi	C 495	18	1.3	174129	2	AC092617	AC092617 Homo sapi
C 423	18	1.3	162361	2	AC023779	AC023779 Homo sapi	C 496	18	1.3	174165	33	AL391480	AL391480 Homo sapi
C 424	18	1.3	162609	9	AC021171	AC021171 Homo sapi	C 497	18	1.3	174465	2	AC022582	AC022582 Homo sapi
C 425	18	1.3	162868	9	AC021034	AC021034 Homo sapi	C 498	18	1.3	174471	2	AC015454	AC015454 Homo sapi
C 426	18	1.3	163147	2	AC024445	AC024445 Homo sapi	C 499	18	1.3	174771	2	AC079150	AC079150 Homo sapi
C 427	18	1.3	163345	2	AL360002	AL360002 Homo sapi	C 500	18	1.3	174832	9	AF280107	AF280107 Homo sapi
C 428	18	1.3	163464	2	AC007462	AC007462 Homo sapi	C 501	18	1.3	174832	2	AC027185	AC027185 Homo sapi
C 429	18	1.3	163612	2	AC001794	AC001794 Homo sapi	C 502	18	1.3	175093	2	AC022367	AC022367 Homo sapi
C 430	18	1.3	163624	2	AC009780	AC009780 Homo sapi	C 503	18	1.3	175184	2	AL583808	AL583808 Homo sapi
C 431	18	1.3	164232	2	AC079064	AC079064 Homo sapi	C 504	18	1.3	175332	2	AC087616	AC087616 Homo sapi
C 432	18	1.3	164518	2	AC068210	AC068210 Homo sapi	C 505	18	1.3	175827	2	AC092664	AC092664 Homo sapi
C 433	18	1.3	165130	2	AL596113	AL596113 Homo sapi	C 506	18	1.3	175827	2	AL358535	AL358535 Homo sapi
C 434	18	1.3	165146	9	AL316444	AL316444 Homo sapi	C 507	18	1.3	175902	2	AP000504	AP000504 Homo sapi
C 435	18	1.3	165228	9	CNS01DWS	AL138479 Human chr	C 508	18	1.3	175940	9	HS0152L7	AL109518 Human DNA
C 435	18	1.3	165243	2	AL162408	AL162408 Homo sapi							

509	18	1.3	176051	9	AC023490	AC023490	Homo sapi	582	18	1.3	188043	2	AL500526	AL500526	Homo sapi
C 510	18	1.3	176056	2	AC0933026	AC0933026	Homo sapi	583	18	1.3	188152	2	AC025847	AC025847	Homo sapi
C 511	18	1.3	176132	2	AC060833	AC060833	Homo sapi	584	18	1.3	188180	2	AC068021	AC068021	Homo sapi
C 512	18	1.3	176379	2	AC060833	AC060833	Homo sapi	585	18	1.3	188358	2	AC050922	AC050922	Homo sapi
C 513	18	1.3	176556	2	AC060833	AC060833	Homo sapi	586	18	1.3	188420	2	AC025205	AC025205	Homo sapi
C 514	18	1.3	176639	9	AC016987	AC016987	Homo sapi	587	18	1.3	188966	2	AC020747	AC020747	Homo sapi
C 515	18	1.3	176645	9	AC068712	AC068712	Homo sapi	588	18	1.3	189385	2	CNS06C7Y	CNS06C7Y	Human chr
C 516	18	1.3	176678	2	AC036213	AC036213	Homo sapi	589	18	1.3	189814	2	AC010754	AC010754	Homo sapi
C 517	18	1.3	176714	9	AC005618	AC005618	Homo sapi	590	18	1.3	190000	2	AC006425	AC006425	Homo sapi
C 518	18	1.3	176724	2	AC091882	AC091882	Homo sapi	591	18	1.3	190220	2	AC068343	AC068343	Homo sapi
C 519	18	1.3	176783	2	AL139281	AL139281	Human DNA	592	18	1.3	190277	2	AC079474	AC079474	Homo sapi
C 520	18	1.3	177307	2	AP003524	AP003524	Oryza sat	593	18	1.3	190746	2	AC087191	AC087191	Homo sapi
C 521	18	1.3	177696	2	AL139815	AL139815	Homo sapi	594	18	1.3	190984	2	AC048332	AC048332	Homo sapi
C 522	18	1.3	177716	2	AP001027	AP001027	Homo sapi	595	18	1.3	191141	2	AC069374	AC069374	Homo sapi
C 523	18	1.3	177717	9	AC061958	AC061958	Homo sapi	596	18	1.3	191244	2	AC024222	AC024222	Homo sapi
C 524	18	1.3	177922	2	AC040902	AC040902	Homo sapi	597	18	1.3	191267	2	AC011844	AC011844	Homo sapi
C 525	18	1.3	178373	2	AL357273	AL357273	Homo sapi	598	18	1.3	191676	2	AP003526	AP003526	Homo sapi
C 526	18	1.3	178401	2	AL353770	AL353770	Homo sapi	599	18	1.3	192021	2	AP003526	AP003526	Homo sapi
C 527	18	1.3	178414	2	AC025341	AC025341	Homo sapi	600	18	1.3	192150	2	AP003025	AP003025	Homo sapi
C 528	18	1.3	178585	2	AC092927	AC092927	Homo sapi	601	18	1.3	192718	2	AC090366	AC090366	Homo sapi
C 529	18	1.3	178714	33	AC022925	AC022925	Homo sapi	602	18	1.3	193026	2	AC010158	AC010158	Homo sapi
C 530	18	1.3	178796	2	AC020667	AC020667	Homo sapi	603	18	1.3	193079	2	AC010907	AC010907	Homo sapi
C 531	18	1.3	179026	2	AC018857	AC018857	Homo sapi	604	18	1.3	193131	2	AL390766	AL390766	Homo sapi
C 532	18	1.3	179134	2	AC022817	AC022817	Homo sapi	605	18	1.3	194256	2	AC019158	AC019158	Homo sapi
C 533	18	1.3	179363	9	AP000842	AP000842	Homo sapi	606	18	1.3	194750	2	AC092961	AC092961	Homo sapi
C 534	18	1.3	179396	2	AC0017056	AC0017056	Homo sapi	607	18	1.3	194874	2	AC080090	AC080090	Homo sapi
C 535	18	1.3	179460	2	AC012512	AC012512	Homo sapi	608	18	1.3	195058	2	AP003087	AP003087	Homo sapi
C 536	18	1.3	179556	2	AC018734	AC018734	Homo sapi	609	18	1.3	195121	2	AC021187	AC021187	Homo sapi
C 537	18	1.3	179622	2	AC092519	AC092519	Felis cat	610	18	1.3	195343	2	AC068342	AC068342	Homo sapi
C 538	18	1.3	179833	9	AL162233	AL162233	Human DNA	611	18	1.3	195472	2	AC018398	AC018398	Homo sapi
C 539	18	1.3	179912	9	AL133415	AL133415	Human DNA	612	18	1.3	195913	2	AL357336	AL357336	Homo sapi
C 540	18	1.3	179947	2	AL591853	AL591853	Homo sapi	613	18	1.3	196267	2	AC026603	AC026603	Homo sapi
C 541	18	1.3	179971	2	AC013395	AC013395	Homo sapi	614	18	1.3	196361	9	AC005386	AC005386	ctld_57_1
C 542	18	1.3	180072	2	AC017092	AC017092	Homo sapi	615	18	1.3	196680	9	AC019235	AC019235	Homo sapi
C 543	18	1.3	180129	2	AC021026	AC021026	Homo sapi	616	18	1.3	196832	9	AC018694	AC018694	Homo sapi
C 544	18	1.3	180499	3	AC007817	AC007817	Drosophila	617	18	1.3	197031	2	AC021155	AC021155	Homo sapi
C 545	18	1.3	180865	2	AP003766	AP003766	Oryza sat	618	18	1.3	197547	2	AC013494	AC013494	Homo sapi
C 546	18	1.3	181137	2	AC084303	AC084303	Homo sapi	619	18	1.3	198239	2	AL358073	AL358073	Homo sapi
C 547	18	1.3	181175	2	AL158214	AL158214	Human DNA	620	18	1.3	198978	2	AC011312	AC011312	Homo sapi
C 548	18	1.3	181264	9	AC016962	AC016962	Homo sapi	621	18	1.3	200337	9	AL590608	AL590608	Human DNA
C 549	18	1.3	181305	2	AC013415	AC013415	Homo sapi	622	18	1.3	201027	2	AC015035	AC015035	Homo sapi
C 550	18	1.3	181899	2	AC015904	AC015904	Homo sapi	623	18	1.3	201035	2	AC016744	AC016744	Homo sapi
C 551	18	1.3	182133	2	AC016235	AC016235	Homo sapi	624	18	1.3	203076	2	AC081647	AC081647	Homo sapi
C 552	18	1.3	182268	2	AC015591	AC015591	Homo sapi	625	18	1.3	203111	2	AP001651	AP001651	Mus muscu
C 553	18	1.3	182274	2	AC023435	AC023435	Homo sapi	626	18	1.3	203241	9	AC006282	AC006282	Homo sapi
C 554	18	1.3	182303	9	AL356750	AL356750	Human DNA	627	18	1.3	203405	2	AC024964	AC024964	Homo sapi
C 555	18	1.3	182304	2	AC092625	AC092625	Homo sapi	628	18	1.3	203407	2	AC006174	AC006174	Homo sapi
C 556	18	1.3	182389	9	AP002829	AP002829	Homo sapi	629	18	1.3	204154	2	AL450352	AL450352	Homo sapi
C 557	18	1.3	182475	2	AC023487	AC023487	Homo sapi	630	18	1.3	204220	2	AC022938	AC022938	Homo sapi
C 558	18	1.3	182512	2	AC027552	AC027552	Homo sapi	631	18	1.3	204310	2	AC017094	AC017094	Homo sapi
C 559	18	1.3	182756	8	AC007789	AC007789	Oryza sat	632	18	1.3	204310	2	AC004387	AC004387	Homo sapi
C 560	18	1.3	182818	2	AP001202	AP001202	Homo sapi	633	18	1.3	206056	2	AC013698	AC013698	Homo sapi
C 561	18	1.3	182833	2	AC011085	AC011085	Homo sapi	634	18	1.3	206388	2	AC021698	AC021698	Homo sapi
C 562	18	1.3	182873	2	AC068044	AC068044	Homo sapi	635	18	1.3	206457	9	AC013268	AC013268	Homo sapi
C 563	18	1.3	183080	2	AC027002	AC027002	Homo sapi	636	18	1.3	206925	2	AC080008	AC080008	Homo sapi
C 564	18	1.3	183279	2	AC079014	AC079014	Homo sapi	637	18	1.3	208326	2	AC022888	AC022888	Homo sapi
C 565	18	1.3	183428	2	AC013733	AC013733	Homo sapi	638	18	1.3	208506	2	AP003530	AP003530	Homo sapi
C 566	18	1.3	183542	9	AL353766	AL353766	Human DNA	639	18	1.3	208765	2	AC010736	AC010736	Homo sapi
C 567	18	1.3	183549	9	AL445473	AL445473	Human DNA	640	18	1.3	210202	2	AC008778	AC008778	Homo sapi
C 568	18	1.3	183549	9	AL161646	AL161646	Human DNA	641	18	1.3	210814	2	HS0543C6	HS0543C6	Homo sapi
C 569	18	1.3	183833	9	AC017022	AC017022	Homo sapi	642	18	1.3	211025	2	AL580875	AL580875	Homo sapi
C 570	18	1.3	183970	2	AC017022	AC017022	Homo sapi	643	18	1.3	212110	2	AL581707	AL581707	Homo sapi
C 571	18	1.3	184010	2	AC018656	AC018656	Homo sapi	644	18	1.3	212656	9	AC007957	AC007957	Homo sapi
C 572	18	1.3	184226	2	AL157786	AL157786	Homo sapi	645	18	1.3	212656	9	AC007957	AC007957	Homo sapi
C 573	18	1.3	185229	2	AP001173	AP001173	Homo sapi	646	18	1.3	214222	2	AC074140	AC074140	Homo sapi
C 574	18	1.3	185664	10	AC006507	AC006507	Mus muscu	647	18	1.3	214445	2	AL365497	AL365497	Homo sapi
C 575	18	1.3	185755	2	AC092146	AC092146	Homo sapi	648	18	1.3	216431	2	AC027321	AC027321	Homo sapi
C 576	18	1.3	185880	2	AC079377	AC079377	Mus muscu	649	18	1.3	216431	2	AL354771	AL354771	Homo sapi
C 577	18	1.3	185939	2	AP002828	AP002828	Homo sapi	650	18	1.3	218675	2	AC030227	AC030227	Homo sapi
C 578	18	1.3	186414	2	AC012454	AC012454	Homo sapi	651	18	1.3	218894	2	AC073152	AC073152	Homo sapi
C 579	18	1.3	186572	2	AC079130	AC079130	Mus muscu	652	18	1.3	219564	2	AC073708	AC073708	Homo sapi
C 580	18	1.3	186896	2	AC037486	AC037486	Homo sapi	653	18	1.3	219635	2	AL390207	AL390207	Homo sapi
C 581	18	1.3	187044	2	AL158164	AL158164	Homo sapi	654	18	1.3	221310	2	AL590439	AL590439	Homo sapi
C 581	18	1.3	187445	2	AC024639	AC024639	Homo sapi	654	18	1.3	221437	9	AC034242	AC034242	Homo sapi

C 655	18	1.3	222477	2	AP001847	AP001847 Homo sapi	C 728	17	1.2	959	2	AC082528	AC082528 Giardina 1
C 656	18	1.3	224058	9	AC008513	AC008513 Homo sapi	C 729	17	1.2	967	2	AC067107	AC067107 Giardina 1
C 657	18	1.3	224400	2	AC008758	AC008758 Homo sapi	C 730	17	1.2	981	14	AB018697	AB018697 Human rot
C 658	18	1.3	224614	3	AE003761	AE003761 Drosophila	C 731	17	1.2	981	14	GA026376	U26376 Group A rot
C 659	18	1.3	227949	10	AC090431	AC090431 Mus muscu	C 732	17	1.2	984	14	AY002485	AY002485 HIV-1 iso
C 660	18	1.3	228239	2	AC011405	AC011405 Homo sapi	C 733	17	1.2	995	2	AC066417	AC066417 Giardina 1
C 661	18	1.3	235586	2	AC024173	AC024173 Mus muscu	C 734	17	1.2	996	2	AC036912	AC036912 Giardina 1
C 662	18	1.3	240825	6	AX087869	AX087869 Sequence	C 735	17	1.2	1001	2	AC054114	AC054114 Giardina 1
C 663	18	1.3	254871	2	AL586136	AL586136 Mus muscu	C 736	17	1.2	1001	9	AE279968	AE279968 Homo sapi
C 664	18	1.3	257728	2	AC006846	AC006846 Caenorhab	C 737	17	1.2	1001	9	AE279969	AE279969 Homo sapi
C 665	18	1.3	258369	2	AC091771	AC091771 Mus muscu	C 738	17	1.2	1001	9	AE279970	AE279970 Homo sapi
C 666	18	1.3	258654	2	AC008579	AC008579 Homo sapi	C 739	17	1.2	1001	9	AE279971	AE279971 Homo sapi
C 667	18	1.3	260464	2	AC006800	AC006800 Caenorhab	C 740	17	1.2	1001	9	AE279972	AE279972 Homo sapi
C 668	18	1.3	273658	2	AC079423	AC079423 Mus muscu	C 741	17	1.2	1001	9	AE279973	AE279973 Homo sapi
C 669	18	1.3	281423	2	AL139335	AL139335 Homo sapi	C 742	17	1.2	1001	9	AE279974	AE279974 Homo sapi
C 670	18	1.3	282895	2	AL133478	AL133478 Homo sapi	C 743	17	1.2	1001	9	AE279975	AE279975 Homo sapi
C 671	18	1.3	287571	2	AC092358	AC092358 Homo sapi	C 744	17	1.2	1001	9	AE279976	AE279976 Homo sapi
C 672	18	1.3	294754	2	AC055741	AC055741 Homo sapi	C 745	17	1.2	1001	9	AE279977	AE279977 Homo sapi
C 673	18	1.3	297201	2	AC091956	AC091956 Homo sapi	C 746	17	1.2	1001	9	AE279978	AE279978 Homo sapi
C 674	18	1.3	299050	1	AP003131	AP003131 Staphyloc	C 747	17	1.2	1001	9	AE279979	AE279979 Homo sapi
C 675	18	1.3	299719	2	AC006858	AC006858 Caenorhab	C 748	17	1.2	1001	9	AE279980	AE279980 Homo sapi
C 676	18	1.3	300829	2	AL359532	AL359532 Homo sapi	C 749	17	1.2	1001	9	AE279981	AE279981 Homo sapi
C 677	18	1.3	303750	1	AP003133	AP003133 Staphyloc	C 750	17	1.2	1001	9	AE279982	AE279982 Homo sapi
C 678	18	1.3	305662	2	AC006746	AC006746 Caenorhab	C 751	17	1.2	1001	9	AE279984	AE279984 Homo sapi
C 679	18	1.3	319012	3	AE003556	AE003556 Drosophila	C 752	17	1.2	1001	9	AE279985	AE279985 Homo sapi
C 680	18	1.3	322774	3	CEV73F8A	AL132862 Caenorhab	C 753	17	1.2	1001	9	AE279986	AE279986 Homo sapi
C 681	18	1.3	329861	1	NMA522491	AL162756 Neisseria	C 754	17	1.2	1001	9	AE279987	AE279987 Homo sapi
C 682	18	1.3	340000	9	AP001688	AP001688 Homo sapi	C 755	17	1.2	1001	9	AE279988	AE279988 Homo sapi
C 683	18	1.3	340000	9	AP001700	AP001700 Homo sapi	C 756	17	1.2	1001	9	AE279989	AE279989 Homo sapi
C 684	18	1.3	343590	1	AP003359	AP003359 Staphyloc	C 757	17	1.2	1001	9	AE279990	AE279990 Homo sapi
C 685	18	1.3	347225	1	AP003351	AP003351 Staphyloc	C 758	17	1.2	1001	9	AE279992	AE279992 Homo sapi
C 686	18	1.3	349600	6	AX044032	AX044032 Sequence	C 759	17	1.2	1001	9	AE279993	AE279993 Homo sapi
C 687	17	1.2	235	10	MM009498	U09498 Mus muscu	C 760	17	1.2	1001	9	AE279994	AE279994 Homo sapi
C 688	17	1.2	285	10	AY011829	AY011829 Dipodomys	C 761	17	1.2	1001	9	AE279995	AE279995 Homo sapi
C 689	17	1.2	294	6	AX112157	AX112157 Sequence	C 762	17	1.2	1001	9	AE279996	AE279996 Homo sapi
C 690	17	1.2	304	11	DM106D9T	Z83507 D. melanoga	C 763	17	1.2	1001	9	AE279997	AE279997 Homo sapi
C 691	17	1.2	341	3	AF021066	AF021066 Achenomor	C 764	17	1.2	1001	9	AE279998	AE279998 Homo sapi
C 692	17	1.2	376	11	G48787	G48787 SHGC-83643	C 765	17	1.2	1001	9	AE279999	AE279999 Homo sapi
C 693	17	1.2	381	5	AF148421	AF148421 Pagrus au	C 766	17	1.2	1001	9	AE280001	AE280001 Homo sapi
C 694	17	1.2	399	14	RGR288006	AJ288006 Human gro	C 767	17	1.2	1001	9	AE280002	AE280002 Homo sapi
C 695	17	1.2	488	11	G49513	V004752 SHGC-68699	C 768	17	1.2	1001	9	AE280003	AE280003 Homo sapi
C 696	17	1.2	498	4	HGGL02	G49513 SHGC-68699	C 769	17	1.2	1001	9	AE280004	AE280004 Homo sapi
C 697	17	1.2	517	9	HSPA23C5	279131 H. sapiens f	C 770	17	1.2	1001	9	AE280005	AE280005 Homo sapi
C 698	17	1.2	629	10	AF303453	AF303453 Mus muscu	C 771	17	1.2	1001	9	AE279983	AE279983 Homo sapi
C 699	17	1.2	631	10	AF303454	AF303454 Mus muscu	C 772	17	1.2	1002	9	AE279983	AE279983 Homo sapi
C 700	17	1.2	638	8	CNAJ2958	AJ002958 Cicer ari	C 773	17	1.2	1002	9	AE279991	AE279991 Homo sapi
C 701	17	1.2	635	10	AY011332	AY011332 Muscardin	C 774	17	1.2	1002	9	AE280000	AE280000 Homo sapi
C 702	17	1.2	653	14	RVGA041	Z80274 Rotavirus g	C 775	17	1.2	1002	9	AE280007	AE280007 Homo sapi
C 703	17	1.2	673	10	AF303451	AF303451 Mus muscu	C 776	17	1.2	1002	9	AE280008	AE280008 Homo sapi
C 704	17	1.2	687	11	G48874	G48874 SHGC-77728	C 777	17	1.2	1014	2	AC057137	AC057137 Giardina 1
C 705	17	1.2	699	5	TAU77178	U77178 Triturus al	C 778	17	1.2	1047	14	PRVC95X	L24165 Porcine rot
C 706	17	1.2	732	6	AX122913	AX122913 Sequence	C 779	17	1.2	1059	8	AB045607	L24164 Porcine rot
C 707	17	1.2	733	9	HSAT5594	AJ005594 Homo sapi	C 780	17	1.2	1059	8	AB045607	AB045607 Homo sapi
C 708	17	1.2	784	10	AF303450	AF303450 Mus muscu	C 781	17	1.2	1062	14	KOBYP7A	AB0261 Bovine rota
C 709	17	1.2	805	1	AF179592	AF179592 Vibrio ch	C 782	17	1.2	1083	3	AB001513	AB001513 Carabus m
C 710	17	1.2	816	6	AB0952	AB0952 Sequence 4	C 783	17	1.2	1083	3	AB001514	AB001514 Carabus m
C 711	17	1.2	816	6	AB95331	AB95331 Sequence 4	C 784	17	1.2	1083	3	AB050677	AB050677 Aristocar
C 712	17	1.2	834	6	AB5688	AB5688 Sequence 34	C 785	17	1.2	1083	3	AB050687	AB050687 Aristocar
C 713	17	1.2	834	6	AR155181	AR155181 Sequence	C 786	17	1.2	1142	5	AF261254	AF261254 Mantidact
C 714	17	1.2	834	6	E65706	E65706 Genome DNA	C 787	17	1.2	1152	4	AF095465	AF095465 Oryctolag
C 715	17	1.2	842	8	AB024274	AB024274 Citrus un	C 788	17	1.2	1158	10	RNLCKTP	Z15029 R. norvegicu
C 716	17	1.2	854	8	AY027357	AY027357 Arabidops	C 789	17	1.2	1341	6	AX143113	AX143113 Sequence
C 717	17	1.2	855	6	AB5702	AB5702 Sequence 36	C 790	17	1.2	1409	14	FLANURAMA	L25815 Influenza v
C 718	17	1.2	855	6	AR155195	AR155195 Sequence	C 791	17	1.2	1409	14	FLANURAMB	L25817 Influenza v
C 719	17	1.2	863	6	E65720	E65720 Genome DNA	C 792	17	1.2	1409	14	FLANURAMC	L25817 Influenza v
C 720	17	1.2	863	6	AB5889	AB5889 Sequence 54	C 793	17	1.2	1413	14	FLANMK	J02146 Influenza A
C 721	17	1.2	863	6	AR155382	AR155382 Sequence	C 794	17	1.2	1433	3	ECOPCNLEU	L07866 Escherichia
C 722	17	1.2	863	6	E65907	E65907 Genome DNA	C 795	17	1.2	1431	8	AF170870	AF170870 Saccharom
C 723	17	1.2	911	2	AC053308	AC053308 Giardia 1	C 796	17	1.2	1458	9	AF177201	AF177201 Homo sapi
C 724	17	1.2	915	13	AF265273	AF265273 Unculture	C 797	17	1.2	1459	10	UBAJ3455	UBAJ3455 Unidentifi
C 725	17	1.2	917	8	AF094426	AF094426 Oxyphils	C 798	17	1.2	1488	10	AF244361	AF244361 Mus muscu
C 726	17	1.2	924	2	AC078530	AC078530 Giardina 1	C 799	17	1.2	1490	1	AF182102	AF182102 Escherich
C 727	17	1.2	938	2	AC065874	AC065874 Giardina 1	C 800	17	1.2	1490	1	AF182104	AF182104 Escherich

C 801	17	1.2	1490	1	AF182105	AF182105 Escherich	874	17	1.2	5642	2	AC091129	AC091129 Mus muscu
C 802	17	1.2	1490	1	AF182106	AF182106 Escherich	875	17	1.2	5678	2	AC017339	AC017339 Drosophila
C 803	17	1.2	1490	1	AF182107	AF182107 Escherich	876	17	1.2	5694	3	AC006687	AC006687 Caenorhab
C 804	17	1.2	1498	1	AF182103	AF182103 Escherich	C 877	17	1.2	5775	5	AB034701	AB034701 Xenopus l
C 805	17	1.2	1507	6	A43711	A43711 Sequence 8	878	17	1.2	5874	9	AF117907	AF117907 Homo sapi
C 806	17	1.2	1507	6	AR106145	AR106145 Sequence	879	17	1.2	5895	17	AF242211	AF242211 Escherich
C 807	17	1.2	1539	10	AF217173	AF217173 Mus muscu	C 880	17	1.2	6329	2	AC197778	AC197778 Drosophila
C 808	17	1.2	1593	9	AF142992	AF142992 Homo sapi	881	17	1.2	6394	9	HSM802644	HSM802644 Homo sapi
C 809	17	1.2	1593	9	AF184344	AF184344 Homo sapi	C 882	17	1.2	6447	3	DM040214	DM040214 Drosophila
C 810	17	1.2	1594	9	HSU94703	HSU94703 Homo sapien	C 883	17	1.2	6511	10	AF261919	AF261919 Mus muscu
C 811	17	1.2	1600	1	EC009194	EC009194 Homo sapi	C 884	17	1.2	6693	9	AF172078	AF172078 Homo sapi
C 812	17	1.2	1612	1	ECORROS	D17549 Escherichia	C 885	17	1.2	7299	1	AF242209	AF242209 Escherich
C 813	17	1.2	1642	1	ECOSTGMA38	D13548 Escherichia	C 886	17	1.2	7350	6	AR145630	AR145630 Sequence
C 814	17	1.2	1686	10	MM094700	U94700 Mus musculu	C 887	17	1.2	7350	6	II14735	II14735 Sequence 14
C 815	17	1.2	1705	5	AF128489	AF128489 Calotes v	C 888	17	1.2	7408	10	AF223080	AF223080 Rattus no
C 816	17	1.2	1770	5	XELXWADA	L17888 Xenopus lae	C 889	17	1.2	7679	8	SCYNR071C	SCYNR071C S.cerevisia
C 817	17	1.2	1818	6	AX109977	AX109977 Sequence	C 890	17	1.2	7850	1	AF242208	AF242208 Escherich
C 818	17	1.2	1846	9	AK026831	AK026831 Homo sapi	C 891	17	1.2	7850	1	AF242210	AF242210 Escherich
C 819	17	1.2	1863	9	AK027744	AK027744 Homo sapi	C 892	17	1.2	8462	10	AF060246	AF060246 Mus muscu
C 820	17	1.2	1898	10	MMMYF5	X56182 M.musculus	C 893	17	1.2	8700	8	AF068574	AF068574 Arabidops
C 821	17	1.2	1959	1	AF182950	AF182950 Homo sapi	C 894	17	1.2	8877	1	BACMCPTLP	BACMCPTLP Bacillus su
C 822	17	1.2	2045	1	AF275947	AF275947 Homo sapi	C 895	17	1.2	9010	14	AF178440	AF178440 Tritoma
C 823	17	1.2	2051	5	XL077639	XL077639 Xenopus lae	C 896	17	1.2	9957	9	AL390767	AL390767 Human DNA
C 824	17	1.2	2129	3	AF336057	AF336057 Pristina	C 897	17	1.2	9973	1	AE004295	AE004295 Vibrio ch
C 825	17	1.2	2240	10	MMIAP1B	AB015497 Passiflor	C 898	17	1.2	10008	1	AE001241	AE001241 Treponema
C 826	17	1.2	2295	8	AB015497	AB015497 Passiflor	C 899	17	1.2	10098	1	AE006255	AE006255 Lactococc
C 827	17	1.2	2344	10	AF021836	AF021836 Mus muscu	C 900	17	1.2	10116	1	AE005326	AE005326 Escherich
C 828	17	1.2	2603	9	AK027702	AK027702 Homo sapi	C 901	17	1.2	10268	1	AE005361	AE005361 Escherich
C 829	17	1.2	2631	14	PHVCGAA	X70418 Pepper Huas	C 902	17	1.2	10271	1	AE000922	AE000922 Methanoba
C 830	17	1.2	2646	8	AF141928	AF141928 Pelargonol	C 903	17	1.2	10495	3	EHCP5	EHCP5 Escherich
C 831	17	1.2	2654	8	GMU20213	U20213 Glycine max	C 904	17	1.2	10573	1	AE005502	AE005502 Escherich
C 832	17	1.2	2654	8	SCYBL097W	X25858 S.cerevisia	C 905	17	1.2	10580	1	AE001784	AE001784 Thermocog
C 833	17	1.2	2689	8	LESBPI	X89929 L.esculentu	C 906	17	1.2	10674	14	AF100465	AF100465 Dendque vi
C 834	17	1.2	2729	8	HSM801458	D7103364 Homo sapi	C 907	17	1.2	10674	14	AF100467	AF100467 Dendque vi
C 835	17	1.2	2751	8	D67034	D67034 Muccor javan	C 908	17	1.2	10674	14	AF100468	AF100468 Dendque vi
C 836	17	1.2	2829	1	AF021260	AF021260 Botreila	C 909	17	1.2	10674	14	AF100469	AF100469 Dendque vi
C 837	17	1.2	2928	8	AB027507	AB027507 Arabidops	C 910	17	1.2	10703	14	DENRC3	DENRC3 Dendque vi
C 838	17	1.2	2938	8	BC007291	BC007291 Homo sapi	C 911	17	1.2	11060	1	AE006380	AE006380 Lactococc
C 839	17	1.2	2949	6	AX099364	AX099364 Sequence	C 912	17	1.2	11093	1	U67595	U67595 Methanococc
C 840	17	1.2	3006	6	AB056773	AB056773 Macaca fa	C 913	17	1.2	11115	1	AE007754	AE007754 Clostridi
C 841	17	1.2	3012	8	CORPOGEM	X63384 O.cenit rpic	C 914	17	1.2	11129	1	AE001228	AE001228 Treponema
C 842	17	1.2	3323	8	SCYNR072W	D71687 S.cerevisia	C 915	17	1.2	11457	1	AE000338	AE000338 Escherich
C 843	17	1.2	3386	5	AF103901	AF103901 Parca fla	C 916	17	1.2	11507	1	AE006586	AE006586 Streptoco
C 844	17	1.2	3386	5	AF103901	AF103901 Parca fla	C 917	17	1.2	11507	1	AE006586	AE006586 Streptoco
C 845	17	1.2	3386	6	AF144813	AF144813 Sequence	C 918	17	1.2	11562	1	AE003124	AE003124 Bacillus
C 846	17	1.2	3393	6	HOMCSPB	U03072 Human serin	C 919	17	1.2	11663	1	AE000242	AE000242 Escherich
C 847	17	1.2	3429	5	AF143003	AF143003 Parca fla	C 920	17	1.2	12339	1	AE008029	AE008029 Agrodac
C 848	17	1.2	3499	9	HSU43519	U33519 Human dylstr	C 921	17	1.2	12343	1	U67505	U67505 Methanococc
C 849	17	1.2	3588	9	HSM802481	U157434 Homo sapi	C 922	17	1.2	13083	1	AE005198	AE005198 Escherich
C 850	17	1.2	3850	6	D84439	D84439 Schistosach	C 923	17	1.2	13086	2	AC019563	AC019563 Drosophila
C 851	17	1.2	3949	6	AY033318	AY033318 Unculture	C 924	17	1.2	13316	1	U67602	U67602 Methanococc
C 852	17	1.2	4126	1	AF270497	AF270497 Escherich	C 925	17	1.2	14763	1	AE000734	AE000734 Agulfex a
C 853	17	1.2	4160	6	AX086127	AX086127 Sequence	C 926	17	1.2	14925	10	MMCLCNV14	MMCLCNV14 Mus muscu
C 854	17	1.2	4160	6	HSM801597	AL136624 Homo sapi	C 927	17	1.2	15251	1	D90786	D90786 E.coli geno
C 855	17	1.2	4164	1	MBFDPHC	MB4798 Methanobact	C 928	17	1.2	15401	8	AC023838	AC023838 Arabidops
C 856	17	1.2	4169	1	AFU33282	AFU33282 Anabaena PC	C 929	17	1.2	15401	8	AP000385	AP000385 Arabidops
C 857	17	1.2	4343	8	SCYNL087W	D71363 S.cerevisia	C 930	17	1.2	15892	3	AE001419	AE001419 Plasmodi
C 858	17	1.2	4363	8	NCUI7251	U17251 Neurospora	C 931	17	1.2	16019	1	D90785	D90785 E.coli geno
C 859	17	1.2	4421	9	HSM802480	AL157433 Homo sapi	C 932	17	1.2	17933	6	SCORFSDNA	SCORFSDNA S.pombe c
C 860	17	1.2	4470	3	AF041423	AF041423 Plasmodi	C 933	17	1.2	18510	2	AC017804	AC017804 Drosophila
C 861	17	1.2	4470	3	AF041424	AF041424 Plasmodi	C 934	17	1.2	18633	2	CELT24C12	CELT24C12 Drosophila
C 862	17	1.2	4499	3	AF041422	AF041422 Plasmodi	C 935	17	1.2	19049	2	AC020853	AC020853 Mus muscu
C 863	17	1.2	4622	8	AF288216	AF288216 Oryza sat	C 936	17	1.2	19349	2	AC012969	AC012969 Drosophila
C 864	17	1.2	4646	14	MSSWVRNA	X73883 Wheat spind	C 937	17	1.2	20343	5	GGVATTIG	GGVATTIG Gcarric
C 865	17	1.2	4726	3	AF210316	AF210316 Drosophila	C 938	17	1.2	21110	8	SPCC2386	SPCC2386 S.pombe c
C 866	17	1.2	4755	9	AB040914	AB040914 Homo sapi	C 939	17	1.2	23250	9	AC093016	AC093016 Homo sapi
C 867	17	1.2	4757	6	AR145629	AR145629 Sequence	C 940	17	1.2	23389	6	HSV526F1A	HSV526F1A Human DNA
C 868	17	1.2	4757	6	AR145629	AR145629 Sequence	C 941	17	1.2	23389	8	AC007294	AC007294 Arabidops
C 869	17	1.2	5089	6	LES6378	LES6378 Sequence	C 942	17	1.2	23486	8	AC091079	AC091079 Homo sapi
C 870	17	1.2	5156	6	AR085703	AR085703 Sequence	C 943	17	1.2	24839	9	AP001973	AP001973 Homo sapi
C 871	17	1.2	5401	10	MUSCSDPPOA	M35526 Mouse compl	C 944	17	1.2	24903	3	CELAC7	CELAC7 Caenorhabd
C 872	17	1.2	5403	10	MUSCSDPPOA	M35526 Mouse compl	C 945	17	1.2	26691	3	CET05H10	CET05H10 Caenorhabd
C 873	17	1.2	5542	9	HSU09954	U09954 Human ribos	C 946	17	1.2	27277	3	CELK0563	CELK0563 Caenorhabd

C 947	17	1.2	27332	3	CER11X8	Z73310 Caenorhabdi
948	17	1.2	27719	2	AC012623	AC012623 Homo sapi
949	17	1.2	27935	3	CELC07D10	U13072 Caenorhabdi
950	17	1.2	28409	3	U41012	U41012 Caenorhabdi
C 951	17	1.2	29688	3	CEC09G5	CEC09G5 Caenorhabdi
952	17	1.2	29729	3	CELC13B9	U21309 Caenorhabdi
953	17	1.2	30000	10	RNIGR2	X17012 Rat IGFRII g
C 954	17	1.2	30389	8	SPAC19D5	Z95531 S.pombe chr
955	17	1.2	30562	8	SPAC3A11	Z95260 S.pombe chr
956	17	1.2	30885	1	AE001580	AE001580 Borrelia
957	17	1.2	31857	2	AC020299	AC020299 Drosophila
958	17	1.2	33715	2	AC020075	AC020075 Drosophila
959	17	1.2	34110	2	AC092538	AC092538 Homo sapi
960	17	1.2	34327	2	AC010426	AC010426 Homo sapi
961	17	1.2	34349	3	CELC10C2	U38852 Caenorhabdi
962	17	1.2	34796	3	CEFS3C11	Z75756 Caenorhabdi
963	17	1.2	34874	2	AC015453	AC015453 Homo sapi
C 964	17	1.2	35031	3	AF052833	AF052833 Trypanoso
965	17	1.2	35409	3	CEC44C10	Z69787 Caenorhabdi
C 966	17	1.2	35484	3	CERT26E4	Z81132 Caenorhabdi
967	17	1.2	35579	9	H5359F1	AL023881 Human DNA
C 968	17	1.2	35762	8	AB026553	AB026553 Arabidops
969	17	1.2	36384	9	AC002309	AC002309 Homo sapi
970	17	1.2	36393	9	D87013	D87013 Homo sapien
C 971	17	1.2	36972	9	AF198491	AF198491 Homo sapi
C 972	17	1.2	37010	3	CER628M15	AC084544 Caenorhab
C 973	17	1.2	37170	9	HSE129H9	Z68224 Human DNA s
974	17	1.2	37302	9	D87007	D87007 Homo sapien
975	17	1.2	37329	3	U97009	U97009 Caenorhabdi
C 976	17	1.2	37329	3	U97009	U97009 Caenorhabdi
977	17	1.2	37515	3	CER647P01	AC084663 Caenorhab
978	17	1.2	37628	3	CER31D4	Z92832 Caenorhabdi
C 979	17	1.2	37836	3	AC024695	AC024695 Caenorhab
980	17	1.2	37836	3	AF045639	AF045639 Caenorhab
C 981	17	1.2	38338	6	AF059493	AF059493 Sequence
982	17	1.2	38433	3	U40942	U40942 Caenorhabdi
983	17	1.2	38553	3	CER25K10	Z92796 Caenorhabdi
C 984	17	1.2	38715	9	AC004237	AC004237 Homo sapi
C 985	17	1.2	38950	3	CER19B2	AL001447 Caenorhab
C 986	17	1.2	39215	3	CED1054	Z74030 Caenorhabdi
C 987	17	1.2	39568	2	AC014823	AC014823 Drosophila
C 988	17	1.2	39702	8	AC002111	AC002111 Genomic s
989	17	1.2	39732	7	AF069529	AF069529 Bacterioph
C 990	17	1.2	39781	8	SPC11739	AL031540 S.pombe c
991	17	1.2	39910	9	AP001039	AP001039 Homo sapi
C 992	17	1.2	39944	9	AC011541	AC011541 Homo sapi
C 993	17	1.2	40014	3	CELF48B9	U40955 Caenorhabdi
994	17	1.2	40070	3	DMC87B1	Z98269 Drosophila
C 995	17	1.2	40301	3	U00065	U00065 Caenorhabdi
C 996	17	1.2	40466	3	CELB0496	U58749 Caenorhabdi
C 997	17	1.2	40633	2	AC017139	AC017139 Drosophila
998	17	1.2	41170	9	AP000951	AP000951 Homo sapi
999	17	1.2	41306	9	AP000523	AP000523 Homo sapi
1000	17	1.2	41587	3	AF052832	AF052832 Trypanoso

ALIGNMENTS

RESULT	1					
LOCUS	AX087935	1362 bp	DNA	PAT	17-MAR-2001	
DEFINITION	Sequence 1 from Patent WO0114535.					
ACCESSION	AX087935					
VERSION	AX087935.1	GI:13396913				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
TITLE	1 (bases 1 to 1362)					
	Schwientek, T. and Clausen, H.					
	udp-n-acetylglucosamine: galactose- ₅ (b)1,3- n					

JOURNAL	-acetylglucosamine- ₅ (a)- ₁ (b)1,6- n
FEATURES	-acetylglucosaminyltransferase, cf. c2gnf3
SOURCE	Patent: WO 0114535-A 1 01-MAR-2001;
	Schwientek, Tilo (DK)
	Location/Qualifiers
	1..1362
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	1..1362
	/note="The gene of C2GNF3"
	/codon_start=-1
	/protein_id="CAC34674.1"
	/db_xref="GI:13396914"
	/translation="MKIFKVFKEFTLOOKVFILFLLMLSLIKLVRLPRKDIY
	LVKSLSTSPFVNRRTHYDEFRVYNGSGITFQEPLEKSLERPRDIDLEDD
	VVATSDSDIYQTLRGRQQLVKEKSPFIATSLVYHKAIVELIATIKYHNTY
	CIHYDRAPDFEYVAMNNLKCFSNIFLASKLAVYKHSRLQADLNLSDLKSI
	QMYVYVNLGQDFPLKSNFELVSEKLNANMLLEVYKPNKSLERFTYHEELRVY
	EYVLPYRINISKEAPHNIIQIFVSGAFYLSQAFYKIFNNISYODFPWMSKDYSP
	DEHFWALIRVPGIPGHSRQADVSQSLQKLVKMYNVEGFEYFCGSHLSVCI
	YGAELMLKDKGHWFANKRDSVDPDILIKCLAEKLEEQQDMITLPSKLEFDRMLT
	TTS"
BASE COUNT	410 a 260 c 279 g 413 t
ORIGIN	
Query Match	100.0%; Score 1362; DB 6; Length 1362;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1362;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 atgaagatccaagtgtattttaacataccctacagcagaagtttcattctctt 60
DB	1 ATGAAGATTTCGAATGTTATTTAAACATACCCACGCAAGAAAGTTTCATCTTTT 60
QY	61 ttaaccctatggcgcctctcttcttgaagcttctaattgtagaagcactcttcgcaa 120
DB	61 TTAAACCCTATGGCTCTCTCTTTGTTAAAGCTTTAAATGAGACAGCTTTCCCA 120
QY	121 aaagacattacttggttgaagtaacctgaagtaacctgccttttgaagaacagatac 180
DB	121 AAAGACATTACTTGTTGAGTACACCCCTAAGTACCTCTTTGTAAGAAACGATAC 180
QY	181 actcgttgaagtagaagtagcagatgaagtttaacttctgggtatctatgaacagag 240
DB	181 ACTCGTTAAGTAGAAGTACGATGATGATTAAGTTTTCGGGTATCTAAGACAGAG 240
QY	241 ccttggaaattggaagaagtcctggaaataagaagaaggaacatcttgaagtagat 300
DB	241 CCTTGGAAATTGGAAGAAGTCCTGGAATAAGAAAGGACATTCATTTGAGAGAT 300
QY	301 gatgattgttggaatgacacagatgattgtaacatttgaagctctgaagagttatgct 360
DB	301 GATGATTGTGGAATGACAGATGATGATTTTCAGACTCTTAAGAGTTATGCT 360
QY	361 caaaagcttctcacaagagaggaagaaagcttcccaatagcctattcttggttgcac 420
DB	361 CAAAAGCTTCTCAAAAGAGAGAGAAAGCTTCCCAATAGCCTATTCTTGTTGTCAC 420
QY	421 aaagatgaattatggttgaagagcttaccatgctatatacaacacacaaatattac 480
DB	421 AAAGATGCAATATGTTGGAAGGCTTATCCATGCTATATACACACACAAATTTAC 480
QY	481 tgcattccattatgacgtgaagacacctgataccttcaaaagttgcacatgaattagct 540
DB	481 TGCATTCATTATGATGCTGAAGCACCTGATACCTTCAAAAGTTGCCATGAACATTTACT 540
QY	541 aagtgcttccaatttctcctccaattagaggtctgtagaatatgacacatt 600
DB	541 AAGTGCTTCCAAATTTCTATCTTCCAAATTAAGGCTGTGGAATATGCCACAT 600
QY	601 tccagactcagagctgatttaattgcttgcgaaccttgaagcttccaatcagatgg 660
DB	601 TCCAGACTCAGAGCTGATTTAATTGCTTGCGACCTTCGAAGTCTTCAATCAGTGG 660

QY 661 aaatattatcaactgtgtggcaagaatttccctgaagtaacttgaattgtg 720
 |||||
 Db 661 AAATATGATATCACTGTGTGGCAAGATTTCCCTGAATCAAAATTTGAATTGGG 720
 QY 721 tcaagattcaaaaaactcaatgtggaataatgttgggaagcgtgtaaaccccaaaagt 780
 |||||
 Db 721 TCAGAGTTAAAAAACTCAATGAGCAAAATATGTGGAGACCGGTAAACCCCAACACAGT 780
 QY 781 aaattgaaagattcactaccatcatgaacttagacggtgtccatgaatgtgaa 840
 |||||
 Db 781 AAATTGAAAGATTCTACTACATCATGAACCTTAGAGGGGTCCTTATGATATGTGAAG 840
 QY 841 ctaccataagagaacaacatctccaaagagaaccccccaataacattcagattgtt 900
 |||||
 Db 841 CTACCAATAGAGCAAAACATCTCCAAAGAGACACCCCAACCAATTCAGATATTGTGT 900
 QY 901 ggcagtgctattttgtttaaagcaagcattgttaatatatttcaacaacccatc 960
 |||||
 Db 901 GGCAGTGCTATTTTGTTTAACTCAAGCATTTGTTAAATATTTTCAACAACCTGCATC 960
 QY 961 gtccaaagacttttgcctgtgtcctaaagacatactctcctgataagcactttggcct 1020
 |||||
 Db 961 GTCCAAAGACTTTTTCCTGCTGCTAAAGACACATACCTCTCGATGAGCACTTTTGGGCT 1020
 QY 1021 accttattccgggttccaaagaactcctgggagagatttccagaatcaacccagagattgtct 1080
 |||||
 Db 1021 ACCTTATTCCGGGTTCAGGAATACCTGGGAGATTTCAGATCAACCCAGATGTGTCT 1080
 QY 1081 gactcgcagaagtaagaactgcctgttcaagtggaattactatgaagccttttctatccc 1140
 |||||
 Db 1081 GATCTGCAGAGTAAGACTGCGCTTGTCAAGTGAATTAAGTAAGAGCTTTTTCATTCCTC 1140
 QY 1141 agcttgacggagactcaaccttgaagcgtgtgattttagagagcctcaaatgaagtg 1200
 |||||
 Db 1141 AGTTGACTGGATCTCACTTCGAACCGTGTGATTTAAGAGCTCCAAATTAAGGTGG 1200
 QY 1201 ctatacaagatggaactgtgttcttaataaattgattctaaagtggaacccatctgt 1260
 |||||
 Db 1201 CTATCAAGATGAGCATTTGGTTGCTTAATTAATTGATTCTTAAGGTGAGACCTTATCTTG 1260
 QY 1261 attaatgcttgcagaaagccttgaagacagacagagagactgagcttgccttga 1320
 |||||
 Db 1261 ATTAATGCTTGGCAAGAAAGCTTGAAGACAGACAGAGACTGGATTCATTGGCTCTCA 1320
 QY 1321 gaaaagttattttagatagaanaatctcactaccacatcatga 1362
 |||||
 Db 1321 GAAAGTTATTATGATTAGAAATCTCACTACACATCATATGA 1362
 RESULT 2
 AF132035 3435 bp mRNA PRI 08-APR-2000
 LOCUS Homo sapiens core 2 beta-1,6-N-acetylglucosaminyltransferase 3
 DEFINITION
 ACCESSION AF132035
 VERSION AF132035.1 GI:7527463
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 3435)
 REFERENCES
 AUTHORS Schwientek, T., Yeh, J. C., Levery, S. B., Keck, B., Merix, G., van
 Kessel, A. G., Fukuda, M. and Clausen, H.
 TITLE Control of O-glycan branch formation. Molecular cloning and
 characterization of a novel thymus-associated core 2 beta-
 6-N-acetylglucosaminyltransferase
 J. Biol. Chem. 275 (15), 11106-11113 (2000)
 JOURNAL MEDLINE
 20219156
 2 (bases 1 to 3435)
 REFERENCE Schwientek, T. and Clausen, H.
 AUTHORS
 TITLE Direct Submission

JOURNAL Submitted (27-FEB-1999) School of Dentistry, University of
 Copenhagen, Nørre Alle 20, Copenhagen 2200, Denmark
 FEATURES
 source
 gene
 CDS
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..3435
 /gene="C2GnT3"
 862..2223
 /gene="C2GnT3"
 /codon_start=1
 /product="core 2 beta-1,6-N-acetylglucosaminyltransferase
 3"
 /protein_id="AF063156.1"
 /db_xref="GI:7527464"
 /translation="MKTFKCYFHHHTLOQKVFILFLMTLSLKILANVRRLPRKNDY
 LVEYSLSFVNRNRYTHVADVEYRIVENCSEIAYSLVHDAWELRHAIYQNHNY
 VVMTSDSDIYQTLRGYAKQLVSEKESPEIAYSLVHDAWELRHAIYQNHNY
 CIIYDRKAPDPTFVAMNNLAKFSNFIASKLAVYAHISRLQADLNLSDLLKSSI
 QMKVILCGDPEPLKSNFELVSEKLNKAMLETKVKNPNSLEFEYHHELRVY
 EYKLPTRTNISKEDAPHNIOIFVSAVYFLSQAFVKYIFNNSIVODFPAMSKDTSP
 DEHFMATLIRVPGITPGELSRSDVDSDLOSKEKTLVMKNYEGFEYSCTGSHRSYCI
 YGAHELRLWLIKDSHWFRANKFDSYVDPLILKCLAEKLEBOQRDMITLPSKLEFDRMLT
 TTS"
 BASE COUNT 1032 a 678 c 651 g 1074 t
 ORIGIN
 Query Match 100.0%; Score 1362; DB 9; Length 3435;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgaagatattcaaatgttattttaaataccttaccagacagaagtttcatcctgtt 60
 Db 862 ATGAAGATATTCAAATGTTATTATTAACATACCTACAGCAGAAAGTTTCATCTGTTT 921
 QY 61 ttaacccatggctcctctcttctttaaagcttcaaatgtgagagactcttcggcaa 120
 Db 922 TTAACCCATGAGCTGCTCTCTTGTAAAGCTCTCAATATGAGACGACTTTTCGGCA 981
 QY 121 aaagacattacttggtagagtaactccctaaagtaactgccttcttgaagaacagatac 180
 Db 982 AAAGACATTTACTGTTGTTGAGTACTCCCTAAGTACTGCTGCTTTGTGAAGAAACAGATAC 1041
 QY 181 actcattgttaagatgaagtgcaagtgatgaagtaacttctgggtatctagaagaag 240
 Db 1042 ACTCATGTTAAAGATGAAGTCAAGTATGAAGTAACTGTGGGTATCTATGAACAGGAG 1101
 QY 241 ccttggaaattggaagaagctctggaataagaagaaggaacatcattgacttggagat 300
 Db 1102 CCTTGGAAATTGGAAGAAGTCTGGAATTAAGAAAGGACATCATGACTTGGAGAGAT 1161
 QY 301 gatgatgttggcaagtaaccagtgatgtgacatttcaagactctaagaagttatgct 360
 Db 1162 GATGATGTTGGCAATGCCAGTATGTGCATTATTCAGACTCTAAGAGGTTATGCT 1221
 QY 361 caaagctgtgtcctcaagagagagaagaagcttcccaatagacctattcttggttccac 420
 Db 1222 CAAAAGCTTGTCTCAAGAGAGAGAAAAGCTTCCCAATAGCCTATTCCTTGGTTGCCAC 1281
 QY 421 aaagatgcaatattgttgaagagcttaccatgcttatatacacacagacacaatttac 480
 Db 1282 AAAGATGCAATATATGTTGTTAAAGGCTTATCCATGTGTATATACAAACACACAAATTTTAC 1341
 QY 481 tgcattcatattgattgattgaagcactatattcctcaagtgatgcatgaacaatttagct 540
 Db 1342 TGCATTCATTAATGATGTAAGCACCCTGATACCTTCAAGTTGCCATGAACAATTACTT 1401
 QY 541 aagtgcttccaaatttcttgccttccaaatgaagagcgtgtggaatgatgcccacatt 600
 Db 1402 AAGTGCTTCCAAATTTTTCATGCTTCCAAATATGAGAGGCTGTGGAATATGCCACCAATT 1461
 QY 601 tccagctccagcgtgatttaattgctgtgcgaccttctgaagttcctcaatccagtg 660

```

Db 1462 TCACAGCTCCAGCTGATTAAATGCTTGCGAGCCTTCTAACTCTTCAATCCAGTGG 1521
QY 661 aatatgttcaactgtgtggaagaatttcccccgaatcaatttcaattgtg 720
Db 1522 AATATNGTATCACTGTGTGGGCAAGATTTCCCTCAAGTCAATTTGAAATGTGG 1581
QY 721 tcagagttgaaaaaactcaatggaagaatatgttggagaagggtgaaaccccaacagt 780
Db 1582 TCAGAGTTGAAAAAACTCAATGAGCAAAATATGTTGAGACCGTGAACCCCAAAAGT 1641
QY 781 aaattgaaagattccttaacatcctgaacttaagcgggtgccttaataatgtag 840
Db 1642 AATTTGGAAAGATTCCTTACCTTCATCTGAACTTAGACGGTCCCTTATGAAATGTGAAG 1701
QY 841 ctaccaaataagacaacactctccaaaggaagaccccccaataacattcagatttgt 900
Db 1702 CTACCAATAAGACAAACATCTCCAAAGGAGACACCCCCATACATTCAGATATTGTT 1761
QY 901 ggaagtgctattttgttgaagcagcattgttaatatatttcaaacacacac 960
Db 1762 GGCAGTGCTATTGTTTAACTCAAGCATTTGTTAAATATTTCACAACTCCATC 1821
QY 961 gtccaagacttttgcctggtctaaagaacacatactctcctgattgagcatttggagt 1020
Db 1822 GTTCAAGACTTTTTCCTGCTCTAAAGACACATCTCTCGATGAGCATTGAGGCT 1881
QY 1021 acctgattcgggttccaaagaactcggggagatttccagttccagccagatgtgtc 1080
Db 1882 ACCTTATTCGGGTTCAGGAATACTGGGGAGATTTCCAGATCACCCAGATGTGTC 1941
QY 1081 gactcagaagtaagactcgccttgcgaagtgaattactaagaagcttttctatccc 1140
Db 1942 GATCTGAGAGTAAGACTCCCTTGTCAAGTGAATTACTAAGACGCTTTTCTATCCC 2001
QY 1141 agttgactgatactcaactctgaaagcgtgtgtattttagggagctcacaataaggtg 1200
Db 2002 AGTTGACGTGAGCTCAGCTTCAGACGCTGTGATTATGAGCTCCAAATTAAGTGTG 2061
QY 1201 cttaacaaagatgagcattggttgcataaattgattcctaagttgagccctactctg 1260
Db 2062 CTATATCAAAAGATGACATTGTTGCAATTAATTTGATCTAAGGTGACCTATCTTG 2121
QY 1261 attaaatgcttgcaagaagccttgaagaagacagagagacgtgactcattggcctca 1320
Db 2122 ATTAATAAGCTTGGCAGAAAGCTTGAAGAACAGCAGACAGACATGATCACTTGCCCTCA 2181
QY 1321 gaaaagttattatgataagaatctcaactacacatcatga 1362
Db 2182 GAAAGTTATTATGATAGAAATCTCACACACATCATGA 2223

```

RESULT 3
AC093259/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone RP11-229C3, WORKING DRAFT SEQUENCE,
 6 unordered pieces.
AC093259
VERSION AC093259.1 GI:15193393
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
TITLE DOE Joint Genome Institute.
AUTHORS Sequencing of Human Chromosome 5
 Unpublished
 2 (bases 1 to 184590)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
 Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

```

COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 494352
Center clone name: RP11-229C3
-----
Summary Statistics
Consensus quality: 178980 bases at least Q40
Consensus quality: 182494 bases at least Q30
Consensus quality: 182978 bases at least Q20
Estimated insert size: 183600; agarose-fp estimation
Quality coverage: 8.93 in Q20 bases; agarose-fp estimation
Quality coverage: 8.9 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
  consists of 6 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  *
  * 1 2136: contig of 2136 bp in length
  * 2137 2236: gap of unknown length
  * 2237 5820: contig of 3584 bp in length
  * 5821 5920: gap of unknown length
  * 5921 8425: contig of 2505 bp in length
  * 8426 8525: gap of unknown length
  * 8526 42924: contig of 34399 bp in length
  * 42925 43024: gap of unknown length
  * 43025 98717: contig of 55693 bp in length
  * 98718 98817: gap of unknown length
  * 98818 184590: contig of 85773 bp in length.
  *
  Location/Qualifiers
    source
      1..184590
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="RP11-229C3"
        /clone_1fb="RPC1 human BAC library 11"
BASE COUNT 56307 a 36971 c 36814 g 53966 t 502 others
ORIGIN
Query Match 92.5%; Score 1260; DB 2; Length 184590;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atgaagatatcaaatgattttaaataccctcaagcagaagaatttcatcctgtt 60
Db 22472 ATGAAGATATTTAAATGTTATTTTAAACATACCTACGCAAGAAAGTTTATCTGTT 22413
QY 61 ttaacctatggctcctccttcttgaagccttcaaatgtgaagacacttccgcaa 120
Db 22412 TTAACCTATGGCTCCTCTCTTTGTTAAAGCTTTAAATGTGAGACGACTCTTCGCCAA 22353
QY 121 aaagacattacttctgttgagagactccctaagtlactcgccttctgtgaagaagaataac 180
Db 22352 AAAGACATTTACTGTTGTTGAGTACCTCCCTAAGTACCTGCGCTTTGTGAAGAAACAGATNC 22293
QY 181 actcatgttaagagatcagatcagatgaattactgtctcgggtatctatgaacagagag 240
Db 22292 ACTCATGTTAAGATGAAGTACGATGATGAAGTTAAGTATGATGATGAACAGAGG 22233
QY 241 ccttggaaattggaagaggtctggaataagaagaagaagacatattgacttggagat 300
Db 22232 CCTTGGAAATTGGAAGAGTGTGAAATTAAGAAAGAGGACATATTGACTTGAAGAT 22173
QY 301 gatgatgtgtggaatgacagatgatgtgacattttagacttaagaaggttatgct 360

```

Db 22172 GATGATGTTGGCAATGACCGATGATTGTGACATTTATACAGACTCTAAGAGGTTTACGT 22113
Qy 361 caaaagctgtctcaaaagagagaaagctccccaatagccattcttggttcac 420
Db 22112 CAAAAGCTTGTCTCAAAGSAGSAGAAAAGCTTCCCAATAGCCTATCTTTGGTGTCCAC 22053
Qy 421 aaagatgcaatlatggttgaagagctlatccatgtctatatatacaacagacaaatttac 480
Db 22052 AAAGATGCAATATGTTGTAAGAGGCTTATCCATGCTATATACACACACAAATATTAC 21993
Qy 481 tccatccattatgatcgttaagagccactataccctcaaaagttgcacataaacttact 540
Db 21992 TCCATCCATTATGATCGTAGGACCCCTGATCCTTCAAAGTTGCCATGACAAATTTACT 21993
Qy 541 aagtgcttcccaatatttccattgtctccaaattagagagctgtgaaatgcccacatt 600
Db 21932 AAGTGCTCTCCAAATATTTTCATGCTTCCAAATAGAGGCTGTGAAATATGCCACAT 21873
Qy 601 tccagatccagagctgatttaattgctgtgagaccttctgaacttcccaatccagtg 660
Db 21872 TCCAGACTCCAGGCTGATTTAAATGCTGTGGAACCTTGGAAGCTTCAATCCAGTGG 21813
Qy 661 aaatatgtatcaactgtgtgtggaagaatttccctgaagtcacaaatttgaattggtg 720
Db 21812 AAATATGTTATCACTTGTGTGGCAAGATTTCCCTCAAGTCAAAATTTGAATTTGG 21753
Qy 721 tccagatgaaaaaatacgaatgaggaataatgtgtggaagagtgaaaccccaaaagt 780
Db 21752 TCCAGATTTAAAAAATCTCAATGAGCAATATGTTGGAACCGGTGAACCCCAACAGT 21693
Qy 781 aaattggaagaatttccattaccatcactgaactgaagagtgccctatgaatgtgaag 840
Db 21692 AAATTGGAAGAATTCATCTTACCATCATGAATTAGACGGCTGACCTATGATATGGAAG 21633
Qy 841 ctaccataagagacaacatctccaaaggaagaccccccaataacattagatttgt 900
Db 21632 CTACCAATAGAGCAACAACATCTCCAAAGAGAACCCCCCAATACATTCAGATTTGTT 21573
Qy 901 ggcagtgctattttgtttaaagcaagaattgttaaatatatttcaacaactccatc 960
Db 21572 GGCAGTGCTATTGTTTCCCTGGGCTTAAGTCAAGATTGTTAAATATTTTCAACAAC 21513
Qy 961 gttcaagacttttgcctcgtgtcctaaagacacatctctcctgaagagcatttgag 1020
Db 21512 GTTCAAGACTTTTTCCTGGGCTTAAGTCAAGACTCTCCGATGAGCAGCACTTTGGGCT 21453
Qy 1021 acctgattcgggttcacgaataactctgggaagatttccagatccagccagagatgtct 1080
Db 21452 ACCTTGATTCGGGTTCCAGGAATACCTGGGGAATTTCCAGATCCAGATGATGTCT 21393
Qy 1081 gactcgaagtaagaactcgcctgtcgaagtgaattactatgaagcttttctatccc 1140
Db 21392 GATCTCAGAGTAAGACTGCGCTTGTCAAGTGAATTAATGAAAGGCTTTTCTATCCC 21333
Qy 1141 agttgactgagatcctacactcgaagcgtgttattatgagagctcgaataaagtgtg 1200
Db 21332 AGTTGACTGAGATCTACCTTCGAGCGTGTGATTTATGAGACTCCAGAAATTAAGTGG 21273
Qy 1201 ctatacaaatgagacattggttcttaataaattgattcttaaggltgaacctatctg 1260
Db 21272 CTATCAAAAGATGACATTTGTTGCTAATTAATTTGATTTAGGTGAGACCTATCTGTG 21213
Qy 1261 attaatgtctgacagaagaagcttgaagaacagcagaagagctgagcttgccttgcctca 1320
Db 21212 ATTAAATGCTTGGCAAAAAGCTTGAAGAACAAGAGAGACTGGATCACTTGGCCCTCA 21153
Qy 1321 gaaaagtattatctgatatagaatactcactacacatcatga 1362
Db 21152 GAAAAGTATTATTATGATAGAAATTCACATACACATCATGA 21111

```

RESULT 4
AX087939

```

LOCUS AX087939 31 bp DNA PAT 17-MAR-2001
DEFINITION Sequence 5 from Patent WO0114535.
ACCESSION AX087939
VERSION AX087939.1 GI:13396917
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 31)
AUTHORS Schwientek, T. and Clausen, H.
TITLE Udp-n-acetylglucosamine-6-(a)-r/(glucnac to galnac) -g(b)1,6- n
JOURNAL -acetylglucosaminyltransferase e, c2gnt3
PATENT: WO 0114535-A 5 01-MAR-2001.
Schwientek, Tilo (DK) ; Clausen, Henrik (DK)
FEATURES
source
BASE COUNT 10 a 6 c 7 g 8 t
ORIGIN
Query Match 1.8%; Score 25; DB 6; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 114 tccgcaaaagacattacttgctt 138
Db 7 TCCGCAAAAGACATTTACTTGCTT 31

```

RESULT 5
AX087937/c 21 bp DNA PAT 17-MAR-2001
LOCUS AX087937
DEFINITION Sequence 3 from Patent WO0114535.
ACCESSION AX087937
VERSION AX087937.1 GI:13396915
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 21)
AUTHORS Schwientek, T. and Clausen, H.
TITLE Udp-n-acetylglucosamine: galactose-g(b)1,3- n
JOURNAL -acetylglucosamine-g(a)-r/(glucnac to galnac) -g(b)1,6- n
PATENT: WO 0114535-A 3 01-MAR-2001.
Schwientek, Tilo (DK) ; Clausen, Henrik (DK)
FEATURES
source
BASE COUNT 5 a 7 c 3 g 6 t
ORIGIN
Query Match 1.5%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 6
CNSOLAY6 696 bp mRNA PLN 02-SEP-1999
LOCUS CNSOLAY6
DEFINITION Botrytis cinerea strain T4 cDNA library under conditions of
nitrogen deprivation.

ACCESSION AL113542.1 GI:5828161
 VERSION CDNA library; nitrogen deprivation.
 KEYWORDS Botryotinia fuckeliana.
 SOURCE Botryotinia fuckeliana
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes; Helotiales; Sclerotiniaceae; Botryotinia.
 REFERENCE 1 (bases 1 to 696)
 AUTHORS Bitton, F., Lewis, C., Fortini, D., Pradier, J.M. and Brygoo, Y.
 TITLE Direct Submission
 JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France
 REFERENCE 2 (bases 1 to 696)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage; CP 5706 91057 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
 COMMENT The CDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.
 FEATURES
 source 1. 696
 Location/Qualifiers
 /organism="Botryotinia fuckeliana"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W51C051"
 BASE COUNT 186 a 158 c 164 g 188 t
 ORIGIN
 Query Match 1.5%; Score 21; DB 8; Length 696;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 243 ttgtgaattggaaagagctct 263
 ||||||||||||||||||
 Db 587 ttgtgaattggaaagagctct 607
 RESULT 7
 AC014188 9643 bp DNA HTG 16-NOV-1999
 LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
 AC014188
 AC014188.1 GI:6437147
 HTG: HTGS-PHASE2.
 fruit fly
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 9643)
 AUTHORS Adams, M. and Venter, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10212574 by the submitter. For further information on this sequence e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced.
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source 1. 9643
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 3212 a 1955 c 1666 g 2810 t
 ORIGIN

Query Match 1.5%; Score 21; DB 2; Length 9643;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 520 gtgcacatgacacattagct 540
 ||||||||||||||||||
 Db 5082 gtgcacatgacacattagct 5102
 RESULT 8
 AC023731/c 24650 bp DNA HTG 18-FEB-2000
 LOCUS Drosophila melanogaster clone RP98-33N24, *** SEQUENCING IN PROGRESS ***, 20 unordered pieces.
 DEFINITION AC023731
 AC023731.2 GI:6996819
 VERSION
 KEYWORDS
 HTG: HTGS-PHASE1.
 fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 24650)
 Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K., Bodota, B., Bouck, J., Bowie, S., Brooks, A., Bulay, C., Bunac, C., Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C., David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L., Guetara, M., Harris, K., Hernandez, J., Hodgson, A., Jia, Y., Jones, M., Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondrjewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Owsal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L., Qulles, M., Reiter, D., Rives, M., Sammel, S., Say, J., Scherer, S., Shan, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugand, R., Taber, P., Taylor, T., Vasquez, J., Vilson, R., Vo, Q., Wabshan, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 24650)
 Wootley, K.C.
 Direct Submission
 Submitted (17-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Feb 18, 2000 this sequence version replaced gi:6984305.
 COMMENT
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: DRJ
 Center clone name: RP98-33N24
 Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-terminator Big Dye; 98% of reads
 Assembly program: Phrap; version 0.980611
 Consensus quality: 7321 bases at least Q40
 Consensus quality: 13196 bases at least Q30
 Consensus quality: 16806 bases at least Q20
 Estimated insert size: 19923; sum-of-contigs estimation
 Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 20 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 834: contig of 834 bp in length
835 854: gap of unknown length
855 1694: contig of 840 bp in length
1695 1714: gap of unknown length
1715 2700: contig of 986 bp in length
2701 3985: gap of unknown length
3986 4005: gap of unknown length
4006 5224: contig of 1219 bp in length
5225 5244: gap of unknown length
5245 6444: contig of 1200 bp in length
6445 7897: gap of unknown length
7898 7917: gap of unknown length
7918 9148: contig of 1231 bp in length
9149 9168: gap of unknown length
9169 10197: contig of 1029 bp in length
10198 10217: gap of unknown length
10218 11737: contig of 1500 bp in length
11738 12664: gap of unknown length
12665 12684: gap of unknown length
12685 13429: contig of 745 bp in length
13430 13449: gap of unknown length
13450 14854: contig of 1405 bp in length
14855 14874: gap of unknown length
14875 16285: contig of 1411 bp in length
16286 16305: gap of unknown length
16306 17671: contig of 1366 bp in length
17672 18991: gap of unknown length
18992 19018: contig of 1307 bp in length
19019 20156: gap of unknown length
20157 20176: gap of unknown length
20177 21638: contig of 1462 bp in length
21639 21658: gap of unknown length
21659 22795: contig of 1137 bp in length
22796 22815: gap of unknown length
22816 24650: contig of 1835 bp in length.
  
```

FEATURES
 source 1. 24650
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="RP96-33N24"

BASE COUNT 6800 a 4962 c 5240 g 7211 t 437 others
 ORIGIN

Query Match 1.5%: Score 21; DB 2: Length 24650;
 Best Local Similarity 100.0%; Pred. No. 4.4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 gttgcacgaacaattagct 540
 Db 20732 GTTGCATGACAAATTAGCT 20712

RESULT 9
 AP002084 84011 bp DNA HTG 25-JUL-2000
 LOCUS Homo sapiens chromosome 4 clone 2179J8 map 4q22-q24, *** SEQUENCING
 DEFINITION IN PROGRESS ***; 13 unordered pieces.
 ACCESSION AP002084
 VERSION AP002084.2 GI:9453892
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Homo sapiens DNA, clone:2179L8.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 84011)
 Tsai,S.F.
 Direct Submission
 Submitted (19-MAY-2000) Shih-Feng Tsai, National Yang-Ming
 University, Institute of Genetics, 155 Li-Kong St. Section 2,
 Beitou, Taipei, Taiwan 11221, Republic of China
 (E-mail:lympetsai@ym.edu.tw, URL:http://genome.ym.edu.tw/
 Tel:886-2-28267043, Fax:886-2-28264930)
 On Jul 25, 2000 this sequence version replaced gi:7959372.
 These sequences are draft human sequences, not finished sequences.
 The sequences are unordered pieces. And gaps between the contigs
 of the same clone are represented as 100 N.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 833: contig of 833 bp in length
834 933: gap of 100 bp
934 934 1718: contig of 785 bp in length
1719 1818: gap of 100 bp
1819 6473: contig of 4655 bp in length
6474 6574 12038: contig of 5465 bp in length
6575 12039 12138: gap of 100 bp
12139 16985: contig of 4847 bp in length
16986 17085: gap of 100 bp
17086 22568: contig of 5483 bp in length
22569 22668: gap of 100 bp
22669 28193: contig of 5525 bp in length
28194 28293: gap of 100 bp
28294 35735: contig of 7442 bp in length
35736 35835: gap of 100 bp
35836 46541: contig of 10706 bp in length
46542 46641: gap of 100 bp
46642 53098: contig of 6449 bp in length
53099 53190: gap of 100 bp
53191 61583: contig of 8393 bp in length
61584 61683: gap of 100 bp
61684 73057: contig of 11374 bp in length
73058 73157: gap of 100 bp
73158 84011: contig of 10854 bp in length.
  
```

FEATURES
 source 1. 84011
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4q22-q24"
 /clone="2179L8"

BASE COUNT 26194 a 14705 c 15392 g 26513 t 1207 others
 ORIGIN

Query Match 1.5%: Score 21; DB 2: Length 84011;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1073 atgtctcgtacgtcagagta 1093
 Db 22215 ATGTCTCGATCTGCGAGGTA 22235

RESULT 10
 AC011375 133008 bp DNA HTG 05-MAY-2000
 LOCUS Homo sapiens chromosome 5 clone CTR-119L1, WORKING DRAFT SEQUENCE.
 DEFINITION 10 unordered pieces.
 ACCESSION AC011375

VERSION AC011375.4 GI:7710555
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 133008)
 AUTHORS DOE Joint Genome Institute.
 TITLES Sequencing of Human Chromosome 5
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 133008)
 AUTHORS DOE Joint Genome Institute.
 TITLES Direct Submission
 JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On May 5, 2000 this sequence version replaced gi:6597091.
 COMMENT
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

Project Information
 Center Project Name: 110165, H275
 Center clone name: CIT978SKB_119L1

Summary Statistics
 Consensus quality: 121552 bases at least Q40
 Consensus quality: 127643 bases at least Q30
 Consensus quality: 129058 bases at least Q20
 Estimated insert size: 140000; pulse field gel estimation
 Estimated insert size: 132108; sum-of-contigs estimation
 Quality coverage: 6.28 in Q20 bases; pulse field gel estimation
 Quality coverage: 6.65 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2822: contig of 2822 bp in length
 * 2823 2922: gap of unknown length
 * 2923 5756: contig of 2834 bp in length
 * 5757 5856: gap of unknown length
 * 5857 9903: contig of 4046 bp in length
 * 9903 10002: gap of unknown length
 * 10002 12177: contig of 2175 bp in length
 * 12177 12278: gap of unknown length
 * 12278 19083: contig of 6806 bp in length
 * 19083 19184: gap of unknown length
 * 19184 29949: contig of 10766 bp in length
 * 29949 30049: gap of unknown length
 * 30049 47593: contig of 17544 bp in length
 * 47593 47693: gap of unknown length
 * 47693 70445: contig of 22752 bp in length
 * 70445 70546: gap of unknown length
 * 70546 95942: contig of 25397 bp in length
 * 95942 95943 96042: gap of unknown length
 * 96042 133008: contig of 36966 bp in length.

FEATURES
 Location/Qualifiers
 1. 133008

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTB-119L1"

BASE COUNT 39430 a 25488 c 25495 g 41692 t 903 others
 ORIGIN

Query Match 1.5%; Score 21; DB 2; Length 133008;
 Best Local Similarity 100.0%; Pred. No. 5.1;

Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1243 aaggtgacccattcttgaatt 1263
 Db 20405 AAGGTGACCCATTCTTGATT 20425

RESULT 11

AC073826

LOCUS AC073826 146190 bp DNA HTG 23-JUN-2000
 DEFINITION Mus musculus clone RP23-93K16, WORKING DRAFT SEQUENCE, 35 unordered

ACCESSION

AC073826

VERSION AC073826.1 GI:8810443

KEYWORDS

HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 146190)

AUTHORS

DOE Joint Genome Institute.

TITLE

Sequencing of Mouse

JOURNAL

Unpublished

2 (bases 1 to 146190)

DOE Joint Genome Institute.

Direct Submission

Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

Project Information

Center Project Name: 1772840

Center clone name: RPCI-23_93K16

Summary Statistics

Consensus quality: 118979 bases at least Q40

Consensus quality: 131063 bases at least Q30

Consensus quality: 133821 bases at least Q20

Estimated insert size: 187180; agarose-gel estimation

Estimated insert size: 142790; sum-of-contigs estimation

Quality coverage: 8.66 in Q20 bases; agarose-gel estimation

Quality coverage: 11.35 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 35 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1067: contig of 1067 bp in length

* 1068 1167: gap of unknown length

* 1168 2519: contig of 1352 bp in length

* 2520 2619: gap of unknown length

* 2620 3814: contig of 1195 bp in length

* 3815 3914: gap of unknown length

* 3915 5132: contig of 1218 bp in length

* 5133 5232: gap of unknown length

* 5233 6499: contig of 1267 bp in length

* 6500 6599: gap of unknown length

* 6600 7609: contig of 1010 bp in length

* 7610 7710: gap of unknown length

* 7710 8792: contig of 1083 bp in length

* 8793 8892: gap of unknown length

* 8893 10240: contig of 1348 bp in length

* 10241 10340: gap of unknown length

* 10341 11532: contig of 1192 bp in length

* 11533 11632: gap of unknown length

* 11633 13022: contig of 1390 bp in length

* 13023 13122: gap of unknown length

DEFINITION Human DNA sequence from clone 168L15 on chromosome 6q26-27. Contains part of RPS6KA2 (ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3)), ESTs, STS, GSSs and Cpg islands, complete sequence.

ACCESSION AL022069

VERSION AL022069.1 GI:3256174

KEYWORDS HRC: Cpg Island, RPS6KA2.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Palmer, S.

TITLE Direct Submission

JOURNAL

COMMENT Submitted (24-JUN-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

On Jun 26, 1998 this sequence version replaced gi:2950404.

requests: clonerequests@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

This sequence is the entire insert of clone 168L15. The true right end of clone 427A4 is at 13623 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone cot15 of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr6

168L15 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Plerier de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR:pcvpa2.

location/Qualifiers

1.178098

source

1.178098 "Homo sapiens"

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/map="q26-27"

/clone_1lb="RPCI-1"

/clone="RPI-168L15"

712..1186

repeat_region

/note="MER42 repeat: matches 1..504 of consensus"

1810..1812

variation

/note="clone 427A4; CCG in this entry; substitution"

/replace="cag"

2594..2661

repeat_region

/note="12 repeat: matches 2579..2660 of consensus"

3435..3437

variation

/note="clone 427A4; ATT in this entry; substitution"

/replace="aat"

3703..3916

repeat_region

/note="MER30 repeat: matches 1..230 of consensus"

3830..3832

variation

/note="clone 427A4; GCT in this entry; substitution"

/replace="gat"

3917..4181

misc_feature

/note="match: GSS: Em:B42442"

4418..5138

repeat_region

/note="LIMB3 repeat: matches 5461..6185 of consensus"

5594..5875

repeat_region

/note="LIMC4 repeat: matches 7120..7423 of consensus"

5917..6000

repeat_region

/note="LIMD3 repeat: matches 7345..7432 of consensus"

repeat_region

6124..6452

/note="LIMC4 repeat: matches 7671..7977 of consensus"

6348..6349

variation

/note="clone 427A4; TG in this entry; deletion"

/replace="tatatacag"

10101<6546..6625,8621..8700,12291..12397,15979..16016,18253..18395,26162..26232,27982..28070,46998..47062,57352..57454,65670..65800,68054..68179,84413..84502,86292..86450,93486..93647,96945..97021,98561..98678,102972..103109,104016..104561)

/gene="RPS6KA2"

/note="ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3); match: CDNAS: Em:L07597 Em:M2848 Em:M20188 Em:M20187 Em:M28489 Em:X85106 Em:U08316 Em:M95169 Em:L07598 Em:L07599 Em:D83013 Em:A0007938 Em:AB016869 Em:AB019245 Em:M58340; match: ESTs: Em:N79949 Em:F06080"

/evidence=not-experimental

/product="dj168L15.1 (ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3))"

6546..104561

/gene="RPS6KA2"

10101<6546..6625,8621..8700,12291..12397,15979..16016,18253..18395,26162..26232,27982..28070,46998..47062,57352..57454,65670..65800,68054..68179,84413..84502,86292..86450,93486..93647,96945..97021,98561..98678,102972..103109,104016..104561)

/gene="RPS6KA2"

/note="match: proteins: Sw:P18654 Sw:P51812 Sw:Q15349 Wp:CE18169"

/codon_start=2

/evidence=not-experimental

/product="dj168L15.1 (ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3))"

/protein_id="G1:6065881"

/db_xref="GI:6065881"

/translation="SGPSEIHDRH/GRSESP/LHCEASLP/TEGKLL/LIDFLRGG/DLPRLSKYPMFTEBDPKYVLAIALALPHLSLGIITDIPKRNILDEBGRKTD/FGSKRAIDHDKRAISRCGTEYMADEVYVRNRHROSADMSGVMPEMTGSDPFO/GVDDKRYMLILKAKLSDPQSLGEGASLIRALFKRNPCKRDAGIDGVEIKRHEF/VIDWNLVLRKEIKPPKPAVGRPEDETFHDPKTRKIPDSGVPSPNANHILRGF/SFVASSLIQPSQODLHKVPVHP/IVQOLHGNNTLFDGEIKEDIGVGSVCKRVH/KADTEVAVKILIKSDKRPDEEELIRYQGNHNTILKDYVDGKFLVYLKMGK/ELDIRLROVRFESERSDVLCITITKMDYILHGVGVHRDLKPSNLLYDESSPSI/RVCDPFGAKOLRAGNGLMTPCYTANVAEVLKRGDYDACCIDMSGLITLYMLAGF/TPRANGPDPTPEETLIARISGKVALSGMNDSDAKDVSMKLVHVDPORTAAOV/LKRPVYNREYLSNQLSRDVLHYKGMATYFALNRTQARLEPVLSNLAQRG/MKRLTSYRL"

7032..7132

repeat_region

/note="MIR repeat: matches 79..206 of consensus"

7096..7098

variation

/gene="RPS6KA2"

/note="clone 427A4; TCT in this entry; substitution"

/replace="ttc"

8297..8299

variation

/gene="RPS6KA2"

/note="clone 427A4; ATA in this entry; substitution"

/replace="aca"

8349..8351

variation

/gene="RPS6KA2"

/note="clone 427A4; GTG in this entry; substitution"

/replace="gcg"

8781..8862

repeat_region

/note="41 copies 2 mer gt 72% conserved"

8788..8859

repeat_region

/note="6 copies 12 mer 81% conserved"

8796..8863

repeat_region

/note="17 copies 4 mer tgtg 78% conserved"

8870..9081

repeat_region

/note="106 copies 2 mer gt 60% conserved"

8875..9082

repeat_region

/note="52 copies 4 mer tgtg 60% conserved"

8877..9548

repeat_region

/note="16 copies 42 mer 55% conserved"

```

repeat_region      8937..9128
                    /note="6 copies 32 mer 66% conserved"
repeat_region      8979..9538
                    /note="10 copies 56 mer 57% conserved"
repeat_region      9203..9554
                    /note="11 copies 32 mer 58% conserved"
repeat_region      9205..9554
                    /note="5 copies 70 mer 63% conserved"
repeat_region      9234..9482
                    /note="3 copies 83 mer 71% conserved"
repeat_region      9334..9533
                    /note="50 copies 4 mer 91gt 58% conserved"
repeat_region      9337..9544
                    /note="4 copies 52 mer 69% conserved"
misc_feature        /note="complement(10081..>10297)"
                    /note="match: EST: Em:AA601621"
repeat_region      10710..10867
                    /note="2 copies 79 mer 96% conserved"
misc_feature        /complement(10959..11359)
                    /note="match: GSS: Em:AO614575"
misc_feature        /complement(10962..11370)
                    /note="match: GSS: Em:AO614609"
misc_feature        11361..11794
                    /gene="RPS6KA2"
misc_feature        11933..12432
                    /note="match: GSS: Em:AO268269"
misc_feature        12623..13623
                    /note="match: GSS: Em:AO128333"
misc_feature        12623..13623
                    /gene="RPS6KA2"
repeat_region      13688..14008
                    /note="match: STS: Em:HS427A4T"
misc_feature        14450..14664
                    /note="MER46B repeat: matches 1..233 of consensus"
                    /gene="RPS6KA2"
                    /note="match: GSS: Em:AO036341"
repeat_region      14665..15015
                    /note="LIMB3 repeat: matches 5816..6184 of consensus"
repeat_region      16306..16449
                    /note="2 copies 72 mer 90% conserved"
repeat_region      16591..16980
                    /note="MTLD repeat: matches 84..505 of consensus"
repeat_region      17132..17492
                    /note="THE1B repeat: matches 1..362 of consensus"
repeat_region      20519..20832
                    /note="AluX repeat: matches 3..312 of consensus"
misc_feature        22162..23991
                    /gene="RPS6KA2"
                    /note="match: ESTs: Em:FL1266 Em:F08927"
repeat_region      23013..23202
                    /note="MER20 repeat: matches 1..218 of consensus"
repeat_region      23872..24197
                    /note="AluY repeat: matches 1..311 of consensus"

Query Match      1.5%; Score 21; DB 9; Length 178098;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 930 attgttaataattttca 950
Db 136272 ATTGTTAAATATTTTCAA 136292

RESULT 15
AC009480 188353 bp DNA PRI 09-MAY-2001
DEFINITION Homo sapiens BAC clone RP11-295N18 from 2, complete sequence.
ACCESSION AC009480
VERSION AC009480.4 GI:11181845
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS Sulistion,J.E., and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE
AUTHORS Lamar,B., Druce,K., Laplant,Y., and Gibson,A.
TITLE The sequence of Homo sapiens BAC clone RP11-295N18
JOURNAL Unpublished
JOURNAL 3 (bases 1 to 188353)
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 188353)
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 188353)
REFERENCE
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 16, 2000 this sequence version replaced gi:7630931.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
-----
Summary Statistics
Center Project name: H_NEO295N18
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPc1-11 human BAC library was made from the blood of one male
donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengsen,R.,
Tateno,M., Catanese,J.J., and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-567F11, 200 bp overlap; the
clone sequenced to the right is CTD-2011N5. Actual start of this
clone is at base position 195 of RP11-295N18; actual end is at base
position 188353 of RP11-295N18.
Location/Qualifiers
1..188353

```

FEATURES
Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-295N18"
/clone_id="RPCT-11"
475..812
/rpt_family="MER2_type"
repeat_region
/rpt_family="Alu"
1776..2046
/rpt_family="Alu"
repeat_region
2057..2349
/rpt_family="Alu"
repeat_region
3373..3766
/rpt_family="MALR"
repeat_region
4300..4609
/rpt_family="Alu"
repeat_region
5226..5508
/rpt_family="L1"
repeat_region
5507..6008
/rpt_family="L1"
repeat_region
10521..10838
/rpt_family="L2"
repeat_region
11126..11176
/rpt_family="L2"
repeat_region
11695..11825
/rpt_family="Alu"
repeat_region
11864..12591
/rpt_family="L1"
repeat_region
12894..13111
/rpt_family="MIR"
repeat_region
13773..13957
/rpt_family="L1"
repeat_region
14043..14110
/rpt_family="Mariner"
repeat_region
14263..14475
/rpt_family="L1"
repeat_region
14466..14512
/rpt_family="L1"
repeat_region
14508..14759
/rpt_family="L1"
repeat_region
14760..14900
/rpt_family="L1"
repeat_region
15073..15197
/rpt_family="MER53"
repeat_region
16391..16556
/rpt_family="MIR"
repeat_region
17403..17742
/rpt_family="MER1_type"
repeat_region
18054..18355
/rpt_family="Alu"
repeat_region
20135..20224
/rpt_family="MER1_type"
repeat_region
20258..20581
/rpt_family="MER2_type"
repeat_region
22584..23537
/rpt_family="L1"
repeat_region
23538..24237
/rpt_family="L1"
repeat_region
25743..25952
/rpt_family="MIR"
repeat_region
30092..30204
/rpt_family="CR1"
repeat_region
30213..30368
/rpt_family="MER1_type"
repeat_region
31615..31802
/rpt_family="MIR"
repeat_region
33760..34070
/rpt_family="Alu"
repeat_region
34251..34295
/rpt_family="CR1"
repeat_region
34380..34499
/rpt_family="CR1"
repeat_region
37237..37286

```

```

/rpt_family="MIR"
36439..38599
/rpt_family="MIR"
repeat_region
40806..40879
/rpt_family="CR1"
repeat_region
45154..45465
/rpt_family="Alu"
repeat_region
45703..45795
/rpt_family="L2"
repeat_region
45953..46125
/rpt_family="L2"
repeat_region
47026..47350
/rpt_family="Alu"
repeat_region
47413..47523
/rpt_family="MIR"
repeat_region
47524..48271
/rpt_family="L1"
repeat_region
48272..48387
/rpt_family="MIR"
repeat_region
49668..49707
/rpt_family="L1"
repeat_region
51195..51294
/rpt_family="L2"
repeat_region
51510..51432
/rpt_family="MER105"
repeat_region
52841..52974
/rpt_family="MER1_type"
repeat_region
52864..53002
/rpt_family="MER1_type"
repeat_region
53069..53120
/rpt_family="MIR"
repeat_region
57789..58078
/rpt_family="Alu"
repeat_region
59027..59157
/rpt_family="MER1_type"
repeat_region
59161..59409
/rpt_family="MER2_type"
repeat_region
59475..60200
/rpt_family="MER2_type"
repeat_region
60251..60774
/rpt_family="MER1_type"
repeat_region
60775..62208
/rpt_family="L1"
repeat_region
62214..62658
/rpt_family="MER1_type"
repeat_region
62825..63134

```

Query Match 1.5%; Score 21; DB 9; Length 188353;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

*Qy 1194 aaggtggttataaagttg 1214
 Db 98676 AAGGTGGCTTATCAAGATGG 98696

RESULT 16
 AL592214
 LOCUS AL592214 195277 bp DNA HTG 02-AUG-2001
 DEFINITION Mus musculus chromosome 1 clone RP23-212124, *** SEQUENCING IN
 PROGRESS ***, in unordered pieces.
 ACCESSION AL592214
 VERSION ~ AL592214.3 GI:14588776
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 1 (bases 1 to 195277)
 AUTHORS Sims, S.
 DIRECT SUBMISSION
 TITLE Submitted (31-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone

REQUESTS: clonerequest@sanger.ac.uk
On Jul 3, 2001 this sequence version replaced gi:14586345.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM2124
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 193820 bases at least Q40
Consensus quality: 194186 bases at least Q30
Consensus quality: 194454 bases at least Q20
Insert size: 194977; sum-of-contigs
Insert size: 186154; 8.1% error; agarose-fp
Quality coverage: 8.50x in Q20 bases; sum-of-contigs Quality
coverage: 8.90x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
----- Location/Qualifiers
source
1. 195277
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="RP23-212124"
/clone_11b="RPCT-23"
1. 43565
/note="assembly-fragment:04211"
43666. 93728
/note="assembly-fragment:03891
fragment:chain:1"
93829. 154402
/note="assembly-fragment:00954
fragment:chain:1"
154503. 195277
/note="assembly-fragment:02416
fragment:chain:1
clone_end:SP6
vector:side:right"

BASE COUNT 63064 a 35235 c 35934 g 60739 t 305 others
ORIGIN
Query Match 1.5%; Score 21; DB 2; Length 195277;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 231 tgacagagagccttggaat 251
|||||
Db 159862 TGAACAGAGCCTTGGAAAT 159882

RESULT 17
AC021626 196788 bp DNA HTG 12-MAR-2000
LOCUS Homo sapiens clone Rpl1-39112, WORKING DRAFT SEQUENCE, 28 unordered
DEFINITION
pieces.
AC021626
AC021626.3 GI:7229839
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 196788)
AUTHORS
Biren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE
Homo sapiens, clone Rpl1-39112

JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 196788)
Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Choepel, Y., Colangelo, M., Collins, S., Collumore, A., Cooke, P.,
Darellano, K., Dewar, K., Domino, M., Doyle, M., Fenster, J.,
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardner, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Lieu, C., Karats, A., Klein, J.,
Landers, J., Leoccky, J., Levine, R., McGuck, A., McKernan, K.,
Macdonald, P., Margulis, N., McManis, P., Morrow, J., Naylor, J.,
Mpheters, R., Meldrum, J., Meneses, L., Morrow, J., Peterson, K.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Poirer, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.

Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6978227.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 15883
Center clone name: 391-1-2
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181094 bases at least Q40
Consensus quality: 189581 bases at least Q30
Consensus quality: 192532 bases at least Q20
Insert size: 206000; agarose-fp
Insert size: 194088; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1281: contig of 1281 bp in length
* 1282 1381: gap of 100 bp
* 1382 2099: contig of 1518 bp in length
* 2900 2999: gap of 100 bp
* 3000 5681: contig of 2682 bp in length
* 5682 5781: gap of 100 bp
* 5782 8351: contig of 2570 bp in length
* 8352 8451: gap of 100 bp
* 8452 10639: contig of 2188 bp in length
* 10640 10739: gap of 100 bp
* 10740 13653: contig of 2916 bp in length
* 13656 13753: gap of 100 bp
* 13756 15927: contig of 2172 bp in length
* 15928 16027: gap of 100 bp
* 16028 19028: contig of 3001 bp in length
* 19029 19128: gap of 100 bp
* 19129 22669: contig of 3541 bp in length
* 22670 22769: gap of 100 bp
* 22770 26791: contig of 4022 bp in length

```

* 26792 26891: gap of 100 bp
* 26892 30845: contig of 3954 bp in length
* 30846 30945: gap of 100 bp
* 30946 35594: contig of 4649 bp in length
* 35595 35694: gap of 100 bp
* 35695 40148: contig of 4454 bp in length
* 40149 40248: gap of 100 bp
* 40249 45042: contig of 4794 bp in length
* 45043 45142: gap of 100 bp
* 45143 49477: contig of 4335 bp in length
* 49478 49577: gap of 100 bp
* 49578 54518: contig of 4941 bp in length
* 54519 54618: gap of 100 bp
* 54619 61982: contig of 7364 bp in length
* 61983 62082: gap of 100 bp
* 62083 67811: contig of 5729 bp in length
* 67812 67911: gap of 100 bp
* 67912 74070: contig of 6159 bp in length
* 74071 74170: gap of 100 bp
* 74171 80685: contig of 6515 bp in length
* 80686 80785: gap of 100 bp
* 80786 88056: contig of 8271 bp in length
* 88057 89136: gap of 100 bp
* 89137 99104: contig of 9948 bp in length
* 99105 99204: gap of 100 bp
* 99205 109014: contig of 9810 bp in length
* 109015 109114: gap of 100 bp
* 109115 122471: contig of 13357 bp in length
* 122472 122571: gap of 100 bp
* 122572 138753: contig of 16182 bp in length
* 138754 138853: gap of 100 bp
* 138854 156691: contig of 17838 bp in length
* 156692 156791: gap of 100 bp
* 156792 173620: contig of 16829 bp in length
* 173621 173720: gap of 100 bp
* 173721 196788: contig of 23068 bp in length.

```

```

FEATURES
    source
        1..196788
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_id="RPCT-11 Human Male BAC"

```

```

misc_feature
    1..1281
        /note="assembly-fragment"
misc_feature
    1382..2899
        /note="assembly-fragment"
misc_feature
    3000..5681
        /note="assembly-fragment"
misc_feature
    5782..8351
        /note="assembly-fragment"
misc_feature
    8452..10639
        /note="assembly-fragment"
misc_feature
    10740..13655
        /note="assembly-fragment"
misc_feature
    13756..15927
        /note="assembly-fragment"
misc_feature
    vector_end:77
        /note="assembly-fragment"
misc_feature
    16028..19028
        /note="assembly-fragment"
misc_feature
    19129..22669
        /note="assembly-fragment"
misc_feature
    22770..26791
        /note="assembly-fragment"
misc_feature
    26892..30845
        /note="assembly-fragment"
misc_feature
    30946..35594
        /note="assembly-fragment"
misc_feature
    35695..40148
        /note="assembly-fragment"
misc_feature
    40249..45042
        /note="assembly-fragment"
misc_feature
    45143..49477
        /note="assembly-fragment"

```

```

misc_feature
    49578..54518
        /note="assembly-fragment"
misc_feature
    54619..61982
        /note="assembly-fragment"
misc_feature
    62083..67811
        /note="assembly-fragment"
misc_feature
    67912..74070
        /note="assembly-fragment"
misc_feature
    74171..80685
        /note="assembly-fragment"
misc_feature
    80786..89056
        /note="assembly-fragment"
misc_feature
    89137..99104
        /note="assembly-fragment"
misc_feature
    99205..109014
        /note="assembly-fragment"
misc_feature
    109115..122471
        /note="assembly-fragment"
misc_feature
    122572..138753
        /note="assembly-fragment"
misc_feature
    138854..156691
        /note="assembly-fragment"
misc_feature
    156792..173620
        /note="assembly-fragment"
misc_feature
    173721..196788
        /note="assembly-fragment"

```

```

BASE COUNT  53672 a 43305 c 43403 g 53703 t 2705 others
ORIGIN
Query Match      1.5%; Score 21; DB 2; Length 196788;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  597 cattccagactccagactga 617
Db  37778 CATTCCAGACTCCAGCTGA 37798

```

```

RESULT  18
LOCUS  AE003487/c 301051 bp DNA INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386053 section 4
ACCESSION  AE003487 AE002593
VERSION  AE003487.1 GI:7292680
KEYWORDS  HTG.
SOURCE  fruit fly.
ORGANISM  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Mscophora; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 301051)
Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Ananidis,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Vandal,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,I.H., Blazek,R.G., Chame,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
Miklos,G.L., Abiri,J.F., Agbayani,A., An,H.J., Basu,A.,
Andrews-Pfannkuch,C., Baldwin,D., Bailew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,U., Beasley,E.M., Beeson,K.Y.,
Berens,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Borchers,M.R., Bouck,J., Brokstein,P., Brottier,P., Burris,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,I.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,

```

Gelbart,W.M., Glasser,K., Glodex,A., Gong,F., Gorrell,J.H., Gu,Z.,
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
 Hernandez,J.R., Houck,J., Hostin,D., Housh,F., Karpén,G.H., Ke,Z.,
 Mei,M.H., Ibegwan,C., Jallali,M., Kalush,F., Karpén,G.H., Ke,Z.,
 Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
 Kravitz,S., Kuip,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
 McLeod,M.P., McPherson,D., Merkulov,G., Mishina,N.V., Mobarly,C.,
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
 Nusken,D.R., Pacle,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
 Pollard,J., Puri,V., Reese,M.C., Reinert,K., Remington,K.,
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
 Stapleton,M., Strong,R., Sun,E., Svitskas,R., Tector,C., Turner,R.,
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
 Weistbach,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 2 (bases 1 to 301051)
 20196006
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 FEATURES
 source
 Location/Qualifiers
 1. 301051
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="X"
 complement(join(359, .801,880, .1071))
 /gene="CG2444"
 /product="Ct8084"
 /db_xref="FLYBASE:FBan0002444"
 /db_xref="FLYBASE:FBgn0030326"
 complement(<359, .>1071)
 /gene="CG2444"
 /map="10D2-10D5"
 /db_xref="FLYBASE:FBan0002444"
 /db_xref="FLYBASE:FBgn0030326"
 complement(join(566, .801,880, .988))
 /gene="CG2444"
 /note="CG2444 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0002444"
 /db_xref="FLYBASE:FBgn0030326"
 /protein_id="AAF48078.1"
 /db_xref="GI:7292681"
 /translation="MSQSTVAFLILGLIIVLGGVGAQAAVAKYKLNSSHPKCYL
 DNTLISPEGLADPLPCVRAECHADGIVFKTCDAAVPPGCKORDFVINRREPA
 CCEKRYNCKKH"
 complement(join(2598, .2650,2696, .2754,2797, .3762,
 3828, .4252,4362, .4911,5320, .6027))
 /gene="CG2448"
 /product="Ct8094"
 /db_xref="FLYBASE:FBan0002448"
 /db_xref="FLYBASE:FBgn0030327"
 complement(<2598, .>6027)
 /gene="CG2448"
 /map="10D3-10D5"
 /db_xref="FLYBASE:FBan0002448"
 /db_xref="FLYBASE:FBgn0030327"
 complement(join(3123, .3762,3828, .4252,4362, .4911,
 5320, .5564))
 /gene="CG2448"
 /note="CG2448 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0002448"
 /db_xref="FLYBASE:FBgn0030327"
 /db_xref="FLYBASE:FBgn0030327"
 /gene="CG2446"
 /product="Ct8088"
 /db_xref="FLYBASE:FBan0002446"
 /db_xref="FLYBASE:FBgn0030328"
 complement(join(7627, .8483,8886, .9805,9872, .9925,
 10135, .>101123))
 /gene="CG2446"
 /map="10D4-10D5"
 /db_xref="FLYBASE:FBan0002446"
 /db_xref="FLYBASE:FBgn0030328"
 complement(<7627, .>11123)
 /gene="CG2446"
 /note="CG2446 gene product [alt 2]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0002446"
 /db_xref="FLYBASE:FBgn0030328"
 /protein_id="AAF48081.1"
 /db_xref="GI:7292684"
 /translation="MSNGKATVSEFETGSTRQPEYCYQLYPOYLKAKRCKKPOEL
 IRDQWONLEPKLIKARGDAMHYDELYOSKMKVOSRKFYOLSYLVKTPVAV
 IOETKARFKLPNLEQALITLNLKGVGTMAEALAAAPDSAPPMADCELAIPFI
 EGIDYTKKEYLNFINHQATVERLNAGVGGDTGDPHSPHYELALMSHYANDSPML
 DDMPPGSGASTGTGSLSTNGNSKYLDGDDTDGVDLDDSGAGRNATSEET
 ENENTNPAATPIQSGEAKNNAAYGAALDDGNSYNSDSTQEPITIDNCTQGT
 ATSTEDGEPITLIDIGIGSGSTPLASDESQDEAPKTNLPILTIPOHSSQNOQ
 KQSPQPHKTNNSITNGQAPALAEBAVTAPOAPSKATAAANGNGNGVLAGED
 EDVAEDEDEDELEEDNEALEADENSNGIVRSKTDQOLAAANAVAVSVAAAGA
 DSAFAIGKRTALHCDMELNAGAGVGVEKSPDKLISE"
 complement(join(7627, .8483,8886, .9681))
 /gene="CG2446"
 /note="CG2446 gene product [alt 1]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0002446"
 /db_xref="FLYBASE:FBgn0030328"
 /protein_id="AAG22344.1"
 /db_xref="GI:10728194"
 /translation="MSNGKATVSEFETGSTRQPEYCYQLYPOYLKAKRCKKPOEL
 IRDQWONLEPKLIKARGDAMHYDELYOSKMKVOSRKFYOLSYLVKTPVAV
 IOETKARFKLPNLEQALITLNLKGVGTMAEALAAAPDSAPPMADCELAIPFI
 EGIDYTKKEYLNFINHQATVERLNAGVGGDTGDPHSPHYELALMSHYANDSPML
 DDMPPGSGASTGTGSLSTNGNSKYLDGDDTDGVDLDDSGAGRNATSEET
 ENENTNPAATPIQSGEAKNNAAYGAALDDGNSYNSDSTQEPITIDNCTQGT
 ATSTEDGEPITLIDIGIGSGSTPLASDESQDEAPKTNLPILTIPOHSSQNOQ
 KQSPQPHKTNNSITNGQAPALAEBAVTAPOAPSKATAAANGNGNGVLAGED
 EDVAEDEDEDELEEDNEALEADENSNGIVRSKTDQOLAAANAVAVSVAAAGA
 DSAFAIGKRTALHCDMELNAGAGVGVEKSPDKLISE"
 complement(join(7627, .8483,8886, .9681))
 /gene="CG2446"


```

/note="CG2446 gene product [alt 3]"
/codon_start=1
/db_xref="FLYBASE:FBan0002446"
/db_xref="FLYBASE:FBgn0030328"
/protein_id="AAF48080.1"
/db_xref="GI:7292683"
/translation="MSNGKATVFFENGSTKPEYCYOLYPOVLKAEKKRCKPQEL
IRLDQWYNEELKIKARKDAHVAHYELVQSMKQSRGKFEYPLSYKVNTPRAY
IOETKARFKIPNLEQATTAALSNKGVOTPAASLLAANAPDSAPMADCIAMIPRI
EGIDITKEILNFVNIQATVERLNAEYGTSPHRLALMSHVADLSPEKL
DMPGPGSGASTGTGSLTNGNSKVLDDGTNDGVVDDDESGAGGRTVATESET
ENENTNPAALPLQSGEAKNNAAVGALQDSDNFVNSDSTSEPIIDNDGTTQT
ATTSTEDDEPIALDIGIGSGSTPLASDSNDEAPKTSILPILPTGSSNQNO
KQSPQPKNTNSTNNGCPAPLAEEVTAPOPASKATAPANGNGNGVYGDG
EDAEDEDEDELDEEDNEALEADESSNGIYRDSKLOOLANKAVDAVSFYAAGA
DSAPIGOKRTALHCDMELKNAGVGVCGRKSPDKLRSE"
join(13155..13381,14485..14887,14949..16011)
/gene="CG1837"
/product="CT5608"
/db_xref="FLYBASE:FBan0001837"
/db_xref="FLYBASE:FBgn0030329"
<13155..>16011
/gene="CG1837"
/map="10D4-10D6"
/db_xref="FLYBASE:FBan0001837"

```

```

Query Match
Best Local Similarity 1.5%; Score 21; DB 3; Length 301051;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 520 gttgcattgaacaattagct 540
|||||
Db 221393 GTTGCATTGAACAATTAGCT 221373

```

```

RESULT 19
LOCUS AX087938 32 bp DNA PAT 17-MAR-2001
DEFINITION Sequence 4 from Patent WO0114535.
ACCESSION AX087938
VERSION AX087938.1 GI:13396916
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 32)
AUTHORS Schwientek,T. and Clausen,H.
TITLE Udp-n-acetylglucosamine: galactose-g(b)1,3-n
-acetylglucosamine-g(a)-r/(glucac to galnac)-g(b)1,6-n
patent: WO 0114535-A 4 01-MAR-2001;
JOURNAL Schwientek, Tilo (DK); Clausen, Henrik (DK)
FEATURES
Source location/Qualifiers
1..32
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="a primer named TSHC99"
BASE COUNT 13 a 4 c 7 g 8 t
ORIGIN

```

```

Query Match
Best Local Similarity 1.5%; Score 20; DB 6; Length 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaagatatcaaatgtta 20
|||||
Db 13 ATGAAGATATCAAAATGTTA 32.

```

```

RESULT 20
HUMPSABD/c HUMPSABD 173 bp DNA PRI 27-APR-1993
LOCUS

```

```

DEFINITION Human protein S beta pseudogene, exon n+3.
ACCESSION M57314.102919
VERSION M57314.1 GI:190562
KEYWORDS protein S; pseudogene; sex hormone binding globulin-like.
SOURCE Human DNA.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 173)
AUTHORS Edehnamdt,C.M., Lundwall,A., Wydro,R. and Stenflo,J.
TITLE Molecular analysis of the gene for vitamin K dependent protein S
and its pseudogene. Cloning and partial gene organization
JOURNAL Biochemistry 29, 7861-7868 (1990)
MEDLINE
FEATURES
Source location/Qualifiers
1..173
/organism="Homo sapiens"
/db_xref="taxon:9606"
exon 3..171
BASE COUNT 55 a 26 c 42 g 50 t
ORIGIN Map position 3p11.3q11.2.

```

```

Query Match
Best Local Similarity 1.5%; Score 20; DB 9; Length 173;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 901 ggcagtcctattgttt 920
|||||
Db 105 GGCAGTCTATTGTTT 86

```

```

RESULT 21
LOCUS HUMPSB11/c 185 bp DNA PRI 08-JAN-1995
DEFINITION Human protein S pseudogene beta (PS-beta), exon 11.
ACCESSION M36575.102918
VERSION M36575.1 GI:190319
KEYWORDS S protein.
SEGMENT 11 of 12
SOURCE Human DNA.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 185)
AUTHORS Ploos van Amstel,H.K., Reltsma,P.H., van der Logt,C.P. and
Berlina,R.M.
TITLE Intron-exon organization of the active human protein S gene PS
alpha and its pseudogene PS beta: duplication and silencing during
primate evolution
JOURNAL Biochemistry 29 (34), 7853-7861 (1990)
MEDLINE
COMMENT Draft entry and computer-readable sequence for [Biochemistry 29,
7853-7861 (1990)] kindly submitted
by H.K.Ploos van Amstel, 13-JUL-1990.
FEATURES
Source location/Qualifiers
1..185
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="3p21-cen"
prim_transcript <1..>185

```

```

/gene="PROS2"
/note="protein S pseudogene beta mRNA and introns"
<1..10
/gene="PROS2"
/note="intron 11"
<11..179
/gene="PROS2"
/note="protein S pseudogene beta, 1; G00-120-757"
/pseudo
/number=11
11..179
/exon
/gene="PROS2"

```

/note="Protein S pseudogene beta, exon 11; G00-120-757;
putative"
180. >185
/gene="PROS2"
/note="Intron 12"

BASE COUNT 59 a 26 c 44 g 56 t
ORIGIN

Query Match 1.5%; Score 20; DB 9; Length 185;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 901 ggcagtcctatttggttt 920
DB 113 ggcagtcctatttggttt 94

RESULT 22
HMPSPS08/c 478 bp DNA PRI 10-JAN-1995

LOCUS HMPSPS08
DEFINITION Human protein S-beta pseudogene, exon 12.
ACCESSION M57861.1 J02917
VERSION M57861.1 GI:190682
KEYWORDS protein S; pseudogene.
SEGMENT 8 of 11
SOURCE Human liver DNA.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 478)
AUTHORS Schmidt, D.K., Taitiro, A.V., Phelps, L.G., Tomczak, J.A. and Long, G.L.
TITLE Organization of the human protein S genes
JOURNAL Biochemistry 29 (34), 7845-7852 (1990)
MEDLINE 9108444

FEATURES
source Location/Qualifiers

1..478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
/map="3p21-cen"
order(M57860.1:345..532.1..95)
/gene="PS-beta"
/number=11
96..264
/gene="PROS2"
/note="G00-120-757"
/number=12
BASE COUNT 151 a 71 c 78 g 178 t
ORIGIN

Query Match 1.5%; Score 20; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 ggcagtcctatttggttt 920
DB 198 ggcagtcctatttggttt 179

RESULT 23
MUSCPCP32 732 bp DNA ROD 27-APR-1993
LOCUS MUSCPCP32
DEFINITION Mouse cytochrome c pseudogene MC3.
ACCESSION K03126
VERSION K03126.1 GI:192872
KEYWORDS cytochrome; cytochrome c; pseudogene.
SEGMENT 2 of 2
SOURCE Mouse (BALB/c) 12 day old embryo DNA (Library of P. Leder), clone MC3.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 732)
AUTHORS Limbach, K.J. and Wu, R.
TITLE Characterization of a mouse somatic cytochrome c gene and three cytochrome pseudogenes
JOURNAL Nucleic Acids Res. 13, 617-630 (1985)
MEDLINE 85215501
FEATURES
source Location/Qualifiers

1..732
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 268 a 136 c 160 g 168 t
ORIGIN About 310 bp after segment 1.

Query Match 1.5%; Score 20; DB 10; Length 732;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ctggaataagaagaagga 281
DB 352 CTGGAATAGAGAGAGGA 371

RESULT 24
FLBHAZD/c 1086 bp ss-RNA VRL 02-AUG-1993

LOCUS FLBHAZD
DEFINITION Influenza B/India/3/89 hemagglutinin (HA) gene (segment 4, HA1 domain), 5' end.
ACCESSION M65168
VERSION M65168.1 GI:325155
KEYWORDS haemagglutinin; receptor-binding protein; transmembrane protein.
SOURCE Influenza virus type B (individual isolate B/India/3/89) RNA.
ORGANISM Viruses; ssRNA negative-strand viruses; Orthomyxoviridae; Influenza virus A and B group.

REFERENCE 1 (bases 1 to 1086)
AUTHORS Rota, P.A., Hemphill, M., Whistler, T., Regnery, H.L. and Kendal, A.
TITLE Antigenic and genetic characterization of the haemagglutinins of recent circulating strains of Influenza B virus
JOURNAL J. Gen. Virol. 73, 2737-2742 (1992)
MEDLINE 93019032
FEATURES
source Location/Qualifiers

1..1086
/organism="Influenza B virus"
/isolate="B/India/3/89"
/db_xref="taxon:11520"
/map="segment 4"
1..1086
/gene="HA"
1..1086
/note="HA1/HA2 cleavage site is at position 1086"
/codon_start=1
/product="hemagglutinin"
/protein_id="A043706.1"
/db_xref="GI:325156"
/translation="MKAIIVLVYISNADRICTGSSSPHYVKTATGDEVNVCY
IPVTTPKSHFANLKGKTRGKLCPCCLNCTDLDVALAPKCGITPSAKASILEV
KPVTFGCEPIIMHDKTRKIRQLPNIIRLSTHNVINAEAPGFGYKIGTSASCPN
VTNNGEFAATMAVAAPKNDNNKTATNPLVEAPYICTEGEDQITVWGFNDNETOWK
LYGSKPOKFTSSANGVTTHVSOIGFEPNOAEDGGLPQSGRIYVDVMOVSKGTGI
TYRGITLPCKVWCASGRSVYKISLELIGEDDLHKKYGGKLNKSKRYTGGHAKAIG
NCFIWKTPPLKLANGTKIRPAKIKER"

BASE COUNT 387 a 250 c 224 g 225 t
ORIGIN

Query Match 1.5%; Score 20; DB 14; Length 1086;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 gtgaaccccccaacagtaa 782

```

|||||
Db      627 GTGAACCCCAACAGTAA 608

RESULT 25
LOCUS   EAU59304      2165 bp      DNA      BCT      20-MAR-1997
DEFINITION Enterobacter agglomerans chitinase Chla_Entag (chla) gene, complete ccs.
ACCESSION U59304
VERSION   U59304.1
KEYWORDS  GI:1899047
SOURCE    Pantoea agglomerans.
ORGANISM  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pantoea.
REFERENCE 1 (bases 1 to 2165)
AUTHORS   Chernin,L.S., de la Fuente,L., Sobolev,V., Haran,S., Vorgias,C.E., Oppenheim,A.B. and Chet,I.
TITLE      Molecular cloning, structural analysis, and expression in Escherichia coli of a chitinase gene from Enterobacter agglomerans Appl. Environ. Microbiol. 63 (3), 834-839 (1997)
JOURNAL    97208202
MEDLINE    2 (bases 1 to 2165)
REFERENCE 2 (bases 1 to 2165)
AUTHORS   Chernin,L.S., de la Fuente,L., Sobolev,V., Ismailov,Z., Haran,S., Vorgias,C.E., Oppenheim,A.B. and Chet,I.
TITLE      Direct Submission
JOURNAL    Submitted (27-MAY-1996) Plant Pathology and Microbiology, Hebrew University of Jerusalem, Faculty of Agriculture, Rehovot 76100, Israel

FEATURES
Source    1. 2165
           /organism="Pantoea agglomerans"
           /strain="IC1270"
           /db_xref="taxon:549"
           /gene="chla"
           /EC_number="3.2.1.14"
           /transl_table=1
           /product="Chitinase Chla_Entag"
           /protein_id="AAB49933.1"
           /db_xref="GI:1899048"
           /translation="MRKFNKPLALLIGSLTCSAQAARAGKPSIASGPRKALIVEYD
           OQATYNNIVKTKTAADVSTSMDSGDAGTAKVILDGKEVWSGAEETGSGANPKY
           NGSGRYOMQVALCNADCTASDTQIYATDSSHLPLEPLENNKRPKQDSKAY
           GSTVEGVIGRNTVKTIPQNTFHLIFGIPVCGDGLNDSIKVGESEFALQNSC
           QGRDFEVSVIDPFAAVQKQKGYTAWDDPKGFQGLMAKQARPDLKLPISIGWT
           LSDPEFPMGDKRDRVQKQKGYTAWDDPKGFQGLMAKQARPDLKLPISIGWT
           YVQLMKDLRALMDQLSAETGKRYELTSAIGAKRIDKRYNTAQNMDIIFLMSYDF
           YGAFDLKNGHOTALKRPGKRGHLYGERGOCVAGOGVAPGKTVAAKYGWGTG
           SGYONNNPFTGATGAPYKGTWENGIVYROIANPFTSIDEMOVSIDATAAEAPYKPEST
           GDLPDPDPKRVQAKKYVLDKQGGCFSEHIDANDNDIIMNNTSISGNSAGQ"

BASE COUNT 543 a 534 c 595 g 493 t
ORIGIN

Query Match 1.5%; Score 20; DB 1; Length 2165;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 ttcatcctgttttaacc 67
|||||
Db 1861 TTTCATCCCTGTTTTRACCC 1880

RESULT 26
LOCUS   AF327184S1 2705 bp DNA ROD 11-JAN-2001
DEFINITION Mus musculus inducible costimulatory protein (Icos) gene, exon 1.
ACCESSION AF327184

```

```

VERSION AF327184.1 GI:12083055
KEYWORDS 1 of 2
SEGMENT house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2705)
AUTHORS McAdam,A.J., Greenwald,R.J., Levin,M.A., Chernova,T., Malenkovich,N., Ling,V., Freeman,G.J. and Sharpe,A.H.
TITLE      The inducible costimulatory (ICOS) molecule is critical for CD40-mediated antibody class switching
JOURNAL    Unpublished
REFERENCE 2 (bases 1 to 2705)
AUTHORS   McAdam,A.J., Greenwald,R.J., Levin,M.A., Chernova,T., Malenkovich,N., Ling,V., Freeman,G.J. and Sharpe,A.H.
TITLE      Direct Submission
JOURNAL    Submitted (08-DEC-2000) Adult Oncology, Dana-Farber Cancer Institute, 44 Binney St. Boston, MA 02115, USA

FEATURES
Source    1. 2705
           /organism="Mus musculus"
           /strain="129/Ola"
           /db_xref="taxon:10090"
           <1382..1439
           /gene="Icos"
           /number=1

BASE COUNT 836 a 537 c 446 g 886 t
ORIGIN

Query Match 1.5%; Score 20; DB 10; Length 2705;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 gtgacattatcagactccta 348
|||||
Db 2146 GTGACATTATCAGACTCTA 2165

RESULT 27
LOCUS   AL391815/c 61274 bp DNA PRI 09-APR-2001
DEFINITION Human DNA sequence from clone RP11-223L24 on chromosome 6 contains STSs and GSSs, complete sequence.
ACCESSION AL391815
VERSION   AL391815.6
KEYWORDS  GI:11121081
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 61274)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      Direct Submission
JOURNAL    Submitted (28-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT   requests: clonerequests@sanger.ac.uk
           On Nov 8, 2000 this sequence version replaced gi:10933315.
           During sequence assembly data is compared from overlapping clones.
           Where differences are found these are annotated as variations
           together with a note of the overlapping clone name. Note that the
           variation annotation may not be found in the sequence submission
           corresponding to the overlapping clone, as we submit sequences with
           only a small overlap as described above.
           The following abbreviations are used to associate primary accession
           numbers given in the feature table with their source databases:
           Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
           on the WORMPEP database can be found at
           http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
           was generated from part of bacterial clone contigs of human
           chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
           Group. Further information can be found at

```

http://www.sanger.ac.uk/HGP/Chr6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-223L24. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP11-223L24 is at 1 in this sequence.
 The true right end of clone RP11-223L24 is at 61175 in this sequence.
 The true right end of clone RP11-223L24 is at 46903 in this
 sequence. This sequence was finished as follows unless otherwise
 noted: all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., phred
 quality >= 30); an attempt was made to resolve all sequencing
 problems, such as compressions and repeats; all regions were
 covered by at least one plasmid subclone or more than one M13
 subclone; and the assembly was confirmed by restriction digest.
 RP11-223L24 is from the library RP11-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/dacpac/home.htm
 VECTOR: pBACe3.6

FEATURES

source

1. 61274
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP11-223L24"
 /clone_1fb="RP11-11.1"
 complement(966..1493)
 misc_feature
 /note="match: GSS: Em:AQ677601"
 1429..1571
 repeat_region
 /note="MIR repeat: matches 113..261 of consensus"
 1620..1758
 repeat_region
 /note="MIR repeat: matches 47..198 of consensus"
 complement(2962..3318)
 misc_feature
 /note="match: GSS: Em:AQ133701"
 complement(3212..3661)
 /note="match: GSS: Em:AQ673685"
 3242..3309
 repeat_region
 /note="2 copies 34 mer 91% conserved"
 3345..3849
 misc_feature
 /note="match: GSS: Em:AQ336206"
 4101..4136
 repeat_region
 /note="LIPAS repeat: matches 6108..6143 of consensus"
 4591..4835
 repeat_region
 /note="LIPB3 repeat: matches 5895..6138 of consensus"
 5358..5571
 repeat_region
 /note="MIR repeat: matches 7..262 of consensus"
 complement(5536..5850)
 misc_feature
 /note="match: GSS: Em:AQ776450"
 5945..6051
 repeat_region
 /note="MRSB repeat: matches 1..110 of consensus"
 6283..6504
 repeat_region
 /note="AluJo repeat: matches 76..288 of consensus"
 7050..7471
 repeat_region
 /note="MSTA repeat: matches 1..426 of consensus"
 9416..9861
 misc_feature
 /note="match: GSS: Em:AQ03676"
 10177..10227
 repeat_region
 /note="L2 repeat: matches 2688..2738 of consensus"
 11006..11158
 repeat_region
 /note="LIM1 repeat: matches 6105..6262 of consensus"
 12018..12195
 repeat_region
 /note="MIR repeat: matches 49..252 of consensus"
 12491..12545
 repeat_region
 /note="MRT1 repeat: matches 309..364 of consensus"
 12708..12842
 repeat_region
 /note="MRT11 repeat: matches 17..157 of consensus"
 12811..12857
 repeat_region
 /note="MRT1J repeat: matches 3..51 of consensus"
 13579..13718
 repeat_region
 /note="L2 repeat: matches 2211..2356 of consensus"
 13897..14157
 misc_feature
 /note="match: GSS: Em:AQ564337"
 14320..14496
 repeat_region
 /note="MERSB repeat: matches 2..177 of consensus"

repeat_region
 14569..14645
 /note="L2 repeat: matches 2620..2713 of consensus"
 repeat_region
 15346..15646
 /note="AluJo repeat: matches 1..303 of consensus"
 18563..18722
 repeat_region
 /note="AluJo/FRM repeat: matches 146..307 of consensus"
 18794..19307
 /note="MRT1D repeat: matches 1..505 of consensus"
 complement(19511..19649)
 misc_feature
 /note="match: STS: Em:G07838"
 19523..19645
 /note="match: GSS: Em:A2046652 Em:A2126672"
 complement(19527..19655)
 misc_feature
 /note="match: GSS: Em:A249601"
 complement(19534..19636)
 misc_feature
 /note="match: STS: Em:G08798"
 complement(19534..19632)
 /note="match: GSS: Em:A2103542"
 complement(19541..19636)
 misc_feature
 /note="match: GSS: Em:AQ683084"
 complement(19542..19656)
 /note="match: GSS: Em:A2020257"
 complement(19543..19671)
 /note="match: STS: Em:G12422 Em:G15742"
 complement(19543..19656)
 misc_feature
 /note="match: GSS: Em:A2040015 Em:A2040457"
 19544..19651
 /note="match: STS: Em:G15739"
 complement(19544..19656)
 /note="match: GSS: Em:A2122363 Em:A2271663
 match: STS: Em:L16408"
 19544..19699
 misc_feature
 /note="match: GSS: Em:A2446995"
 19544..19663
 /note="match: GSS: Em:AQ52676 Em:AQ528204"
 19544..19656
 /note="match: GSS: Em:A2318431 Em:A2370392"
 complement(19544..19639)
 /note="match: GSS: Em:B51766"
 19544..19640
 /note="match: STS: Em:G27274"
 complement(19546..19645)
 misc_feature
 /note="match: STS: Em:L17665"
 complement(19546..19658)
 /note="match: STS: Em:G08426"
 19547..19644
 /note="match: GSS: Em:AQ379803"
 19549..19635
 /note="match: STS: Em:G10237"
 complement(19549..19655)
 /note="match: GSS: Em:A2253034"
 complement(19552..19648)
 /note="match: STS: Em:G08887"
 19816..19964
 repeat_region
 /note="MRT1 repeat: matches 1..141 of consensus"
 21610..21675
 repeat_region
 /note="33 copies 2 mer ga 69% conserved"
 21621..21880
 repeat_region
 /note="15 copies 4 mer aaag 83% conserved"
 21738..21859
 repeat_region
 /note="FIAM_A repeat: matches 8..129 of consensus"
 24497..24568
 repeat_region
 /note="18 copies 4 mer ttcc 87% conserved"
 24512..24609
 repeat_region
 /note="49 copies 2 mer ct 65% conserved"
 24656..24953
 repeat_region
 /note="AluJg repeat: matches 1..297 of consensus"
 25114..25425
 repeat_region
 /note="AluJx repeat: matches 1..312 of consensus"
 25493..25726
 /note="MIR repeat: matches 31..261 of consensus"
 25951..26102
 repeat_region
 /note="LIM3 repeat: matches 6004..6162 of consensus"

```

repeat_region      26257..27168
                    /note="LMC5 repeat: matches 6640..7515 of consensus"
repeat_region      27243..27608
                    /note="THEIB repeat: matches 1..364 of consensus"
repeat_region      27609..28708
                    /note="THEIB-INTERNAL repeat: matches 490..1580 of
                    consensus"
repeat_region      28709..29020
                    /note="AluY repeat: matches 1..311 of consensus"
repeat_region      29021..29508
                    /note="THEIB-INTERNAL repeat: matches 1..490 of consensus"
repeat_region      29509..29870
                    /note="THEIB repeat: matches 5..364 of consensus"
repeat_region      29871..30258
                    /note="AluSc repeat: matches 3..274 of consensus"
repeat_region      30678..31184
                    /note="L1MB4 repeat: matches 5795..6287 of consensus"
repeat_region      31378..31489
                    /note="56 copies 2 mer tt 63% conserved"
repeat_region      31492..31676
                    /note="L1MB4 repeat: matches 6007..6180 of consensus"
repeat_region      31773..33606
                    /note="L1MB4 repeat: matches 4425..6303 of consensus"

```

```

Query Match      1.5%: Score 20; DB 9; Length 61274;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      566 ctcccaattagagctgtg 585
        |||||
Db      750 CTTCCAAATTAGAGGCTGTG 731

```

```

RESULT 28
AL513123      70339 bp      DNA      PRI      27-Apr-2001
LOCUS      Human DNA sequence from clone RP11-346D19 on chromosome 6, complete
DEFINITION
ACCESSION      AL513123
VERSION      AL513123.10      GI:13897160
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 70339)
REFERENCE      1
AUTHORS      Blakey/S.
JOURNAL      Direct Submission
Submitted (27-Apr-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1BA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Apr 30, 2001 this sequence version replaced gi:13443468.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP. Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at

```

<http://www.sanger.ac.uk/HGP/Chr6>
 RP11-346D19 is from the library RPCI-11.2 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAE3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-346D19. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP11-367G18 is at 70240 in this
 sequence. The true right end of clone RP1-124C6 is at 100 in this
 sequence.

FEATURES

```

source      location/Qualifiers
            1..70339
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone="RP11-346D19"
            /clone_lib="RPCI-11.2"
            628..921
            /note="AluYb repeat: matches 1..306 of consensus"
            1834..2459
            /note="L2 repeat: matches 1483..2147 of consensus"
            2578..2875
            /note="AluJo repeat: matches 5..298 of consensus"
            3839..3934
            /note="MLT1-INTERNAL repeat: matches 918..1012 of
            consensus"
            4002..4201
            /note="MER5B repeat: matches 2..175 of consensus"
            4929..5186
            /note="AluSg1 repeat: matches 1..287 of consensus"
            5346..5578
            /note="L2 repeat: matches 1660..1887 of consensus"
            5781..5875
            /note="L2 repeat: matches 2171..2268 of consensus"
            5876..6188
            /note="AluY repeat: matches 1..311 of consensus"
            6189..6256
            /note="L2 repeat: matches 2268..2337 of consensus"
            6257..6550
            /note="AluSg repeat: matches 3..300 of consensus"
            6555..6859
            /note="AluSx repeat: matches 1..305 of consensus"
            6860..6971
            /note="L2 repeat: matches 2337..2438 of consensus"
            6988..7152
            /note="L1PB3 repeat: matches 5966..6147 of consensus"
            7567..7748
            /note="L2 repeat: matches 2541..2725 of consensus"
            8539..8842
            /note="AluSc repeat: matches 1..307 of consensus"
            9810..10013
            /note="MLT1 repeat: matches 195..406 of consensus"
            10020..10099
            /note="MIR repeat: matches 1..80 of consensus"
            11344..11550
            /note="L1 repeat: matches 1..80 of consensus"
            12685..13049
            /note="MER47A repeat: matches 5511..5751 of consensus"
            13489..13683
            /note="MER47A repeat: matches 1..366 of consensus"
            13746..13949
            /note="MER47A repeat: matches 67..265 of consensus"
            15394..15701
            /note="AluSg repeat: matches 1..308 of consensus"
            16252..16912
            /note="L2 repeat: matches 1374..2084 of consensus"
            17367..17434
            /note="L2 repeat: matches 2682..2750 of consensus"
            19746..19884
            /note="MLT1 repeat: matches 421..541 of consensus"
            19885..19994
            /note="MER81 repeat: matches 1..114 of consensus"
            19995..20117
            /note="MLT1 repeat: matches 245..421 of consensus"

```

```

repeat_region 20677..20718
/note="21 copies 2 mer aa 76% conserved"
repeat_region 20738..21064
/note="MUT1A1 repeat: matches 1..350 of consensus"
repeat_region 21144..21330
/note="MIR repeat: matches 24..240 of consensus"
repeat_region 21404..21706
/note="AlusX repeat: matches 3..307 of consensus"
repeat_region 22476..22608
/note="L2 repeat: matches 2565..2707 of consensus"
repeat_region 25443..25693
/note="L1MC1 repeat: matches 6067..6321 of consensus"
repeat_region 25897..27083
/note="Tiggr3b repeat: matches 1..1176 of consensus"
repeat_region 27207..27369
/note="MERSA repeat: matches 23..189 of consensus"
repeat_region 27221..27390
/note="MERSA repeat: matches 1..166 of consensus"
repeat_region 27492..27618
/note="MIR repeat: matches 122..261 of consensus"
repeat_region 28470..28713
/note="MIR repeat: matches 2..262 of consensus"
repeat_region 29425..29600
/note="MERS4 repeat: matches 1..176 of consensus"
repeat_region 31817..32196
/note="MSTA repeat: matches 1..426 of consensus"
repeat_region 32198..32245
/note="24 copies 2 mer ac 97% conserved"
repeat_region 32393..32535
/note="FLAM.C repeat: matches 1..143 of consensus"
repeat_region 33417..33714
/note="AluIO repeat: matches 1..296 of consensus"
repeat_region 34739..35171
/note="MUTIF repeat: matches 98..536 of consensus"
repeat_region 35430..35520
/note="AlusG/X repeat: matches 217..307 of consensus"
repeat_region 35544..35845
/note="AluY repeat: matches 1..301 of consensus"
repeat_region 36708..36782
/note="MDEL repeat: matches 1..76 of consensus"
repeat_region 36789..36881
/note="MERSB repeat: matches 67..177 of consensus"
repeat_region 36902..37629
/note="L1MC4 repeat: matches 6969..7648 of consensus"
repeat_region 37634..37685
/note="MERSB repeat: matches 126..178 of consensus"
repeat_region 38202..38300
/note="MERSA repeat: matches 9..106 of consensus"
repeat_region 38469..38656
/note="MERSA repeat: matches 1..184 of consensus"
repeat_region 38669..38782
/note="MERSB repeat: matches 92..178 of consensus"
repeat_region 38800..39708
/note="L1MC1 repeat: matches 2682..3528 of consensus"
repeat_region 39709..40383
/note="L1R8 repeat: matches 1..691 of consensus"
repeat_region 40384..42758
/note="L1ME1 repeat: matches 3528..6105 of consensus"
repeat_region 42845..42931
/note="MERSB repeat: matches 51..133 of consensus"
repeat_region 42960..43209
/note="AluY repeat: matches 1..249 of consensus"
repeat_region 43210..43273
/note="MERSB repeat: matches 2..64 of consensus"
repeat_region 43585..43732
/note="FRAM repeat: matches 1..148 of consensus"
repeat_region 44578..44891
/note="AluDb repeat: matches 1..304 of consensus"
repeat_region 45144..45444
/note="AlusX repeat: matches 1..299 of consensus"
repeat_region 45473..46079
/note="L1PA10 repeat: matches 5546..6160 of consensus"
repeat_region 46406..46483

```

```

repeat_region /note="MERS4 repeat: matches 1..81 of consensus"
48101..48261
/note="MERS1B repeat: matches 2..158 of consensus"
50465..50568
/note="MIR repeat: matches 15..123 of consensus"
50987..51068
/note="MUT1E repeat: matches 50..133 of consensus"
51171..51429
/note="MLT1E repeat: matches 265..557 of consensus"
52002..52259
/note="AlusX repeat: matches 33..290 of consensus"
52819..53599
/note="L1PA2 repeat: matches 2..776 of consensus"
53595..56841
/note="L1PA2 repeat: matches 900..6146 of consensus"
59147..59190

Query Match 1.5%: Score 20; DB 9; Length 70339;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1226 ctaataattgatctaaag 1245
Db 36036 CTAATAATTGATCTAAAG 36055
|||||

RESULT 29
AC005694 82101 bp DNA PRI 01-JUN-2000
LOCUS Homo sapiens chromosome 22q12 clone bk294c2, complete sequence.
DEFINITION AC005694
ACCESSION AC005694.3 GI:496908
VERSION HTG.
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Unpublished
2 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (23-SEP-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (04-DEC-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (19-DEC-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (23-DEC-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (05-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
7 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

TITLE Direct Submission
JOURNAL Submitted (10-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 8 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (12-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 9 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (24-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 10 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (26-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 11 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (28-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 12 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (06-FEB-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 13 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (23-FEB-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 14 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (16-MAR-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 15 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (18-MAR-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 16 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (04-JUN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 17 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (24-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 18 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (26-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 19 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission

JOURNAL Submitted (27-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 20 (bases 1 to 82101)
TITLE Fang, F., Fu, Y., Pan, H. and Roe, B. A.
JOURNAL Direct Submission
Submitted (01-JUN-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Jun 4, 1999 this sequence version replaced gi:4263797.
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below
AC005529(bx256d12) 9641 180736 (137732) overlaps AC005694(bx294c2)
1 82091 (10) AC005527(489d1) 98413 149308 (0) overlaps
AC005694(bx294c2) 1 50869 (3123) AC002991(n47he) 26513 38140 (0)
overlaps AC005694(bx294c2) 1 11605 (70496).
location/Qualifiers
1. 82101
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22q12"
/clone="bx294c2"
FEATURES
SOURCE
BASE COUNT 21863 a 18718 c 19364 g 22156 t
ORIGIN
Query Match 1.5%; Score 20; DB 9; Length 82101;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 711 tgaatgtgtcagaattga 730
Db 63545 TGAATGTGTGACAGTTGA 63564
RESULT 30
AC007164
LOCUS Homo sapiens BAC clone RP11-304A10 from 7p21-p22, complete
DEFINITION
ACCESSION AC007164 GI:7243907
VERSION AC007164.3
KEYWORDS HTG.
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 89862)
TITLE Sulston, J.E. and Waterston, R.
JOURNAL Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE
AUTHORS 2 (bases 1 to 89862)
TITLE Holmes, A., Kallick, J., Pape, K. and Jones, T.
JOURNAL The sequence of Homo sapiens BAC clone RP11-304A10
unpublished
REFERENCE
AUTHORS 3 (bases 1 to 89862)
TITLE Waterston, R.H.
JOURNAL Direct Submission
Submitted (24-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS 4 (bases 1 to 89862)
TITLE Waterston, R.H.
JOURNAL Direct Submission
Submitted (15-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS 5 (bases 1 to 89862)
TITLE Waterston, R.
JOURNAL Direct Submission
Submitted (27-APR-2000) Department of Genetics, Washington

COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 15, 2000 this sequence version replaced gl:4662667.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0304A10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tatenoe, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP4-668E10, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-304A10. Actual end is at base position 62268 of RP4-668E10.

Location/Qualifiers

1. 89862

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7p21-p22"

/clone="RP11-304A10"

/clone_lib="RPc1-11"

340. 457

/rpt_family="MIR"

461. 512

/rpt_family="MIR"

513. 921

/rpt_family="MIR"

922. 1031

/rpt_family="MIR"

2360. 2834

/rpt_family="Retroviral"

2889. 3572

/rpt_family="L2"

3825. 3923

/rpt_family="MIR"

5967. 6042

/rpt_family="MIR"

6530. 6990

repeat_region

repeat_region

repeat_region

6998. 7186

/rpt_family="L1"

repeat_region

7240. 7318

/rpt_family="(TA)n"

repeat_region

7319. 7391

/rpt_family="Alu"

repeat_region

7392. 7749

/rpt_family="L1"

repeat_region

7750. 7945

/rpt_family="L1"

repeat_region

misc.feature

note="match to EST AW003673 (NTD:95850789) ws61d12.x1"

repeat_region

7966. 8085

/rpt_family="MIR"

repeat_region

8086. 8227

/rpt_family="Retroviral"

repeat_region

9431. 9531

/rpt_family="MIR"

repeat_region

9763. 9867

/rpt_family="MIR"

repeat_region

10886. 10907

/rpt_family="AT-rich"

repeat_region

11444. 11549

/rpt_family="Alu"

repeat_region

12509. 12539

/rpt_family="(TTTA)n"

repeat_region

12554. 13483

/rpt_family="L1"

repeat_region

14143. 14638

/rpt_family="L1"

repeat_region

14711. 14882

/rpt_family="MIR"

repeat_region

15988. 16200

/rpt_family="MIR"

repeat_region

16325. 16627

/rpt_family="Alu"

repeat_region

18028. 18081

/rpt_family="(TG)n"

repeat_region

18135. 18384

/rpt_family="MIR"

repeat_region

19044. 19278

/rpt_family="MIR"

repeat_region

19721. 19800

/rpt_family="CT-rich"

repeat_region

/rpt_family="L1"

6998. 7186

/rpt_family="L1"

7240. 7318

/rpt_family="(TA)n"

7319. 7391

/rpt_family="Alu"

7392. 7749

/rpt_family="L1"

7750. 7945

/rpt_family="L1"

7905. 8499

/note="match to EST AW003673 (NTD:95850789) ws61d12.x1"

7966. 8085

/rpt_family="MIR"

8086. 8227

/rpt_family="Retroviral"

9431. 9531

/rpt_family="MIR"

9763. 9867

/rpt_family="MIR"

10886. 10907

/rpt_family="AT-rich"

11444. 11549

/rpt_family="Alu"

12509. 12539

/rpt_family="(TTTA)n"

12554. 13483

/rpt_family="L1"

14143. 14638

/rpt_family="L1"

14711. 14882

/rpt_family="MIR"

15988. 16200

/rpt_family="MIR"

16325. 16627

/rpt_family="Alu"

18028. 18081

/rpt_family="(TG)n"

18135. 18384

/rpt_family="MIR"

19044. 19278

/rpt_family="MIR"

19721. 19800

/rpt_family="CT-rich"

19825. 20135

/rpt_family="L2"

21951. 22034

/rpt_family="L2"

22071. 22364

/rpt_family="Alu"

23130. 23251

/rpt_family="L2"

24315. 24519

/rpt_family="MIR"

25403. 25429

/rpt_family="AT-rich"

27193. 27234

/rpt_family="AT-rich"

27400. 29368

/rpt_family="L1"

29376. 29635

/rpt_family="Alu"

29536. 29626

/rpt_family="(TA)n"

29631. 29753

/rpt_family="L1"

29935. 30283

/rpt_family="L1"

30306. 30385

/rpt_family="Alu"

33392. 33417

/rpt_family="AT-rich"

repeat_region


```

repeat_region      33435..33473
                    /rpt_family="L2"
repeat_region      33481..33655
                    /rpt_family="MER1_type"
repeat_region      33690..33988
                    /rpt_family="Alu"
repeat_region      34079..34402
                    /rpt_family="L1"
repeat_region      34416..34585
                    /rpt_family="(TA)n"
repeat_region      34614..34670
                    /rpt_family="(TA)n"
repeat_region      34956..35480
                    /rpt_family="L2"
repeat_region      35556..35687
                    /rpt_family="MER1_type?"
repeat_region      35742..35873
                    /rpt_family="MER1_group"
repeat_region      36352..36403

```

```

Query Match      1.5%, Score 20; DB 9; Length 89862;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 613 gctgatttaattgcttgc 632
    ||||||||||||||||
Db 77206 GCTGATTAAATGCTTGC 77225

```

```

RESULT 31
AL354692/2c      110443 bp      DNA      PRI      25-JUL-2001
DEFINITION      Human DNA sequence from clone RP11-361M4 on chromosome 9, complete
                  sequence.
ACCESSION      AL354692
VERSION      AL354692.23 GI:15022181
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 110443)
AUTHORS      Wilson,S.
TITLE      Direct Submision
JOURNAL      Submitted (25-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
                  CB10 1SA, UK. E-mail enquiries: humgeny@sanger.ac.uk
                  requests: clonerequests@sanger.ac.uk
                  On Jul 26, 2001 this sequence version replaced gi:15021138.
                  During sequence assembly data is compared from overlapping clones.
                  Where differences are found these are annotated as variations
                  together with a note of the overlapping clone name. Note that the
                  variation annotation may not be found in the sequence submission
                  corresponding to the overlapping clone, as we submit sequences with
                  only a small overlap as described above.
                  This sequence was finished as follows unless otherwise noted: all
                  regions were either double-stranded or sequenced with an alternate
                  chemistry or covered by high quality data (i.e., phred quality >=
                  30); an attempt was made to resolve all sequencing problems, such
                  as compressions and repeats; all regions were covered by at least
                  one plasmid subclone or more than one M13 subclone; and the
                  assembly was confirmed by restriction digest. The following
                  abbreviations are used to associate primary accession numbers given
                  in the feature table with their source databases: Em, EMBL; Sw,
                  SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
                  database can be found at
                  http://www.sanger.ac.uk/Projects/C_elegans/wormpep
                  This sequence was generated from part of bacterial clone compis of human
                  chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
                  Group. Further information can be found at
                  http://www.sanger.ac.uk/HGP/Chr9
                  RP11-361M4 is from the library RPO1-11.2 constructed by the group
                  of Pieter de Jong. For further details see
                  http://www.chori.org/bacpac/home.htm

```

VECTOR: pBAC3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-361M4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-361M4 is at 108444 in this sequence. The true right end of clone RP11-59M22 is at 2000 in this sequence.

```

FEATURES
    source          1..110443
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="9"
                    /clone="RP11-361M4"
                    /clone_1fb="RP11-11.2"
                    1..4068
                    /note="LIP2 repeat: matches 1478..5546 of consensus"
repeat_region      4069..4362
                    /note="AluX repeat: matches 1..293 of consensus"
repeat_region      4363..3588
                    /note="LIP2 repeat: matches 258..1478 of consensus"
                    5566..5949
                    /note="LIM1 repeat: matches 2018..2400 of consensus"
                    5950..6235
                    /note="AluX repeat: matches 1..303 of consensus"
                    6236..9260
                    /note="LIM1 repeat: matches 1390..2018 of consensus"
                    9684..9810
                    /note="LIM4 repeat: matches 5317..5431 of consensus"
                    9835..10361
                    /note="LIP16 repeat: matches 5636..6150 of consensus"
                    10362..10669
                    /note="AluX repeat: matches 1..309 of consensus"
                    10670..11196
                    /note="LIP16 repeat: matches 5084..5636 of consensus"
                    12642..12750
                    /note="MIR repeat: matches 54..168 of consensus"
                    12824..13019
                    /note="L1 repeat: matches 2639..2830 of consensus"
                    13195..13426
                    /note="116 copies 2 mer ta 63% conserved"
                    13449..13617
                    /note="L1 repeat: matches 2296..2464 of consensus"
                    14554..14854
                    /note="AluX repeat: matches 1..303 of consensus"
                    15520..15667
                    /note="MER63A repeat: matches 25..172 of consensus"
                    15863..16146
                    /note="AluB repeat: matches 1..286 of consensus"
                    16205..16533
                    /note="AluX repeat: matches 1..296 of consensus"
                    18551..18875
                    /note="MER69 repeat: matches 1..336 of consensus"
                    18903..19120
                    /note="MER69 repeat: matches 405..629 of consensus"
                    19826..20125
                    /note="AluX repeat: matches 1..301 of consensus"
                    20374..20752
                    /note="MUT1J repeat: matches 5..369 of consensus"
                    20859..22019
                    /note="LMCB repeat: matches 114..1295 of consensus"
                    22022..22088
                    /note="Alu repeat: matches 230..293 of consensus"
                    22092..22257
                    /note="L1 repeat: matches 4789..4975 of consensus"
                    26196..26223
                    /note="14 copies 2 mer tg 100% conserved"
                    27112..27915
                    /note="LIM4 repeat: matches 6004..6211 of consensus"
                    29575..29898
                    /note="MER33 repeat: matches 2..324 of consensus"
                    30358..30404
                    /note="MIR repeat: matches 117..163 of consensus"
                    30500..30567

```

```

repeat_region      /note="MIR repeat: matches 64. .135 of consensus"
32058. .32383
repeat_region      /note="MER33 repeat: matches 2. .324 of consensus"
34237. .34337
repeat_region      /note="MIR repeat: matches 20. .123 of consensus"
34385. .34553
repeat_region      /note="PRAM repeat: matches 4. .172 of consensus"
35130. .35249
repeat_region      /note="MIR repeat: matches 15. .136 of consensus"
35278. .35408
repeat_region      /note="MIR repeat: matches 124. .258 of consensus"
35513. .35591
repeat_region      /note="L2 repeat: matches 2630. .2708 of consensus"
36246. .36463
repeat_region      /note="MER58A repeat: matches 5. .224 of consensus"
36474. .36567
repeat_region      /note="MIR repeat: matches 89. .190 of consensus"
38006. .38530
repeat_region      /note="MLTID repeat: matches 9. .505 of consensus"
39325. .39622
repeat_region      /note="Alusg repeat: matches 1. .239 of consensus"
41143. .41277
repeat_region      /note="MIR repeat: matches 92. .227 of consensus"
41278. .41380
repeat_region      /note="MLTID repeat: matches 402. .505 of consensus"
41417. .41608
repeat_region      /note="MLTID repeat: matches 13. .205 of consensus"
42905. .43159
repeat_region      /note="LTRIC3 repeat: matches 114. .387 of consensus"
44346. .44687
repeat_region      /note="L2 repeat: matches 2361. .2697 of consensus"
44688. .44830
repeat_region      /note="MLTIE repeat: matches 398. .567 of consensus"
46666. .45880
repeat_region      /note="LIM4 repeat: matches 3128. .3374 of consensus"
45928. .45979
repeat_region      /note="LIM4 repeat: matches 2958. .3011 of consensus"
46147. .46446
repeat_region      /note="Alusx repeat: matches 1. .299 of consensus"
46929. .47027
repeat_region      /note="L2 repeat: matches 2607. .2706 of consensus"
47128. .47308
repeat_region      /note="L2 repeat: matches 2290. .2475 of consensus"
47379. .47555
repeat_region      /note="LIM4c repeat: matches 1245. .1416 of consensus"
47584. .48131
repeat_region      /note="LIM4c repeat: matches 1499. .2065 of consensus"
48784. .49492
repeat_region      /note="LIME repeat: matches 2407. .3139 of consensus"
49581. .49870
repeat_region      /note="LIME repeat: matches 3292. .3573 of consensus"
49871. .49985
repeat_region      /note="LIME repeat: matches 7623. .7740 of consensus"
49986. .51673
repeat_region      /note="LIME repeat: matches 3573. .5320 of consensus"
51751. .51944
repeat_region      /note="LIME repeat: matches 5509. .5707 of consensus"
53856. .54072
repeat_region      /note="MER20 repeat: matches 1. .218 of consensus"
55390. .55802
repeat_region      /note="LIR33 repeat: matches 38. .435 of consensus"
57393. .57693
repeat_region      /note="Alusg repeat: matches 1. .302 of consensus"
63187. .63299
repeat_region      /note="L2 repeat: matches 2632. .2750 of consensus"
63321. .64037
repeat_region      /note="L2 repeat: matches 1913. .2710 of consensus"
65180. .66203
repeat_region      /note="L2 repeat: matches 240. .1234 of consensus"
66204. .66477
repeat_region      /note="MLTIA1 repeat: matches 1. .365 of consensus"
66550. .67584
repeat_region      /note="L2 repeat: matches 1431. .2481 of consensus"

```

```

repeat_region      67581. .67651
repeat_region      /note="MIR repeat: matches 142. .212 of consensus"
69580. .69887
repeat_region      /note="AluY repeat: matches 1. .307 of consensus"
71133. .71329
repeat_region      /note="MIR repeat: matches 55. .247 of consensus"
71439. .71603
repeat_region      /note="MIR repeat: matches 5. .181 of consensus"
71633. .71686
repeat_region      /note="L2 repeat: matches 2646. .2703 of consensus"
71960. .72259
repeat_region      /note="AluDb repeat: matches 1. .312 of consensus"

Query Match      1.5%; Score 20; DB 9; Length 110443;
Best Local Similarity 100.0%; Prid.No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1226 ctaataattgatcctaag 1245
Db      61560 CTAATAATTGATCTCTAG 61547

RESULT      32
AC011285      120891 bp      DNA      HTG      31-AUG-2000
LOCUS      Homo sapiens chromosome CTD clone CTD-2337B17, WORKING DRAFT
DEFINITION
ACCESSION      AC011285.5 GI:9954810
VERSION      HTG: HTGS-PHASE1; HTGS-DRAFT.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 120891)
AUTHORS      Waterston,R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 120891)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (05-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT      On Aug 31, 2000 this sequence version replaced g1:9838031.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_MS2337B17
----- Summary Statistics -----
Sequencing vector: M13; 92%
Sequencing vector: Plasmid; 8%
Chemistry: Dye-primer ET; 92% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.99013
Consensus quality: 116247 bases at least Q40
Consensus quality: 120875 bases at least Q30
Consensus quality: 123092 bases at least Q20
Insert size: 92000; agarose-efp
Insert size: 120440; sum-of-configs
Quality coverage: 15.95 in Q20 bases; sum-of-configs
Quality coverage: 15.07 in Q20 bases; sum-of-configs

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

misc_feature /note="assembly_name:Contig122"
21093..22827
/note="assembly_name:Contig123"
misc_feature /note="assembly_name:Contig123"
22928..24125
/note="assembly_name:Contig124"
misc_feature /note="assembly_name:Contig124"
24226..27014
/note="assembly_name:Contig125"
misc_feature /note="assembly_name:Contig125"
27115..28847
/note="assembly_name:Contig126"
misc_feature /note="assembly_name:Contig127"
28948..30686
/note="assembly_name:Contig127"
misc_feature /note="assembly_name:Contig129"
30787..120891
/note="assembly_name:Contig129"
clone_end:sp6
vector_side:right"
BASE COUNT 39086 a 22644 c 22727 g 34516 t 1918 others
ORIGIN
Query Match 1.5%; Score 20; DB 2; Length 120891;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy 7 atattcaatgctatttta 26
|||||
Db 51929 ATATCAAAAGTATTTTAA 51948

RESULT 33
LOCUS HSA118000
DEFINITION Homo sapiens NF2 gene.
ACCESSION Y18000
KEYWORDS Y18000.1 GI:3980299
ALU-like repetitive sequence; L1 repeat; Line repeat; MER repeat;
MIR repeat; NF2 gene; repetitive sequence.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 126138)
Zucman-Rossi,J., Legoux,P., Der Sarkissian,H., Cheret,G., Sor,F.,
Bernardi,A., Cazes,L., Giraud,S., Ollagnon,E., Lenoir,G. and
Thomas,G.
NF2 gene in neurofibromatosis type 2 patients
J. Mol. Genet. 7 (13), 2095-2101 (1996)
MEDLINE 99035680
REFERENCE 2 (bases 1 to 126138)
AUTHORS Legoux,P., Legrand,M.F., Ollagnon,E., Lenoir,G. and
Zucman-Rossi,J.
Characterisation of 16 polymorphic markers in the NF2 gene:
application to hemizyosity detection
Hum. Mutat. 13 (4), 290-293 (1999)
TITLE 3 (bases 1 to 126138)
JOURNAL Zucman-Rossi,J.
MEDLINE 99235548
REFERENCE 3 (bases 1 to 126138)
AUTHORS Zucman-Rossi,J.
Direct Submission
Submitted (25-AUG-1998) J. Zucman-Rossi, INSERM U34/CEPH, 27 rue
Ulfette Dodu, 75010 Paris, FRANCE
JOURNAL Location/Qualifiers
FEATURES
SOURCE 1..126138
/organism="Homo sapiens"
/db_xref="taxon:9606"
repeat_region 31..171
/rpt_family="LIMD2"
/rpt_unit=31..171
/rpt_type=DISPERSD
repeat_region 206..486
/rpt_family="Alu-S2"
/rpt_unit=206..486
/rpt_type=DISPERSD
repeat_region 881..1161
/rpt_family="Alu-Sg"

```

```

/rpt_unit=884..1161
/rpt_type=DISPERSED
repeat_region 1239..1526
/rpt_family="Alu-Sg"
/rpt_unit=1239..1526
/rpt_type=DISPERSED
repeat_region 1564..1848
/rpt_family="Alu-Jo"
/rpt_unit=1564..1848
/rpt_type=DISPERSED
repeat_region 1868..1922
/rpt_family="MUTIE"
/rpt_unit=1868..1922
/rpt_type=DISPERSED
repeat_region 1923..2079
/rpt_family="Alu-Jb"
/rpt_unit=1923..2079
/rpt_type=DISPERSED
LTR 2124..2809
/note="LTR 12"
LTR 2811..2974
/note="LTR 12"
repeat_region 2975..3565
/rpt_family="PER5"
/rpt_unit=2975..3565
/rpt_type=DISPERSED
repeat_region 3574..3704
/rpt_family="Alu-Jb"
/rpt_unit=3574..3704
/rpt_type=DISPERSED
repeat_region 3730..3935
/rpt_family="MER20"
/rpt_unit=3730..3935
/rpt_type=DISPERSED
repeat_region 3981..4117
/rpt_family="MIR"
/rpt_unit=3981..4117
/rpt_type=DISPERSED
repeat_region 4120..4403
/rpt_family="Alu-Sx"
/rpt_unit=4120..4403
/rpt_type=DISPERSED
repeat_region 4425..4489
/rpt_family="MIR"
/rpt_unit=4425..4489
/rpt_type=DISPERSED
repeat_region 4772..4864
/rpt_family="Alu-Sg"
/rpt_unit=4772..4864
/rpt_type=DISPERSED
repeat_region 4969..4988
/rpt_unit=4969..4988
/rpt_type=DISPERSED
repeat_region 4988..5276
/rpt_family="Alu-Y"
/rpt_unit=4988..5276
/rpt_type=DISPERSED
repeat_region 5317..5463
/rpt_family="Alu-Jo"
/rpt_unit=5317..5463
/rpt_type=DISPERSED
repeat_region 5470..5760
/rpt_family="Alu-S2"
/rpt_unit=5470..5760
/rpt_type=DISPERSED
repeat_region 5781..5908
/rpt_family="Alu-Jo"
/rpt_unit=5781..5908
/rpt_type=DISPERSED
repeat_region 6197..6479
/rpt_family="Alu-Y"
/rpt_unit=6197..6479
/rpt_type=DISPERSED
repeat_region 6869..7036
/rpt_family="MIR"
/rpt_unit=6869..7036
/rpt_type=DISPERSED
repeat_region 7063..7351
/rpt_family="Alu-Sg"
/rpt_unit=7063..7351
/rpt_type=DISPERSED
repeat_region 7668..7949
/rpt_family="Alu-Sp"
/rpt_unit=7668..7949
/rpt_type=DISPERSED
repeat_region 7950..7990
/rpt_family="trinucleotide repeat, taa"
/rpt_unit=7950..7990
/rpt_type=DIRECT
variation 8240
/note="polymorphism"
/frequency=".58"
/replace="C"
variation 8348
/note="polymorphism"
/frequency=".54"
/replace="tttttt"
variation 8787
/note="polymorphism"
/frequency=".43"
/replace="a"
exon 8819..9278
/gene="NF2"
8881
/gene="NF2"
/note="polymorphism"
/frequency=".31"
/replace="g"
CDS join(8991..9104,41239..41364,43577..43699,46686..46769,
59142..59210,60079..60161,62673..62748,65689..65823,
69470..69544,72814..72927,76297..76419,77740..77957,
79307..79412,82675..82802,85918..86080,99228..99278)
/gene="NF2"
/note="isoform I"
/codon_start=1
/product="NF2 protein"
/protein_id="CA76992.1"
/db_xref="GI:3980300"
/translation="MAGIASRMSESSLRKRPKPTFTVRIVTMDAEMEFNCENKRGK
DLPLVORTISLRTEWFGLOYTKDVIAMKMKVLDHVSKEEVPYTHLAKFP
ENAEDELVOETQHLFLOYKQKIDDKIYCPASVILASVAQAKGYDPSVHR
GFLAOEELPKRVINLQMPMEKEITANYAHRBRADAEMLTKAOLEMWG
DKIDVFNKSSKLRVNLILQICIGNHDLFMRKADSLEVOQKQAEKARKOM
EEROLAREKOMEAEERTDELRLLOMKEAEAMAEALMSEETADLLAEKQITE
EAKLAKAEAEOMORIKATAIRTEEEKRLQOEVLAEVLAALMAEESERAKE
ADOLKODLOEAREBARKKOLEIATRPYPPNPPIAPLPIDIPFNIGDSLSD
EKDDMKLESMIEIRKVEYEMKSKHLOEOLNEKLEIEMALKERETALDILNENS
DRGSSKNITKKTLDLSAKSRVAFPEL"
join(8991..9104,41239..41364,43577..43699,46686..46769,
59142..59210,60079..60161,62673..62748,65689..65823,
69470..69544,72814..72927,76297..76419,77740..77957,
79307..79412,82675..82802,85918..86080,87501..87611)
/gene="NF2"
/note="isoform II"
/codon_start=1
/product="NF2 protein"
/protein_id="CA76993.1"
/db_xref="GI:3980301"
/translation="MAGIASRMSESSLRKRPKPTFTVRIVTMDAEMEFNCENKRGK
DLFDLCRLGLRTEWFGLOYTKIDIVAMKMKVLDHVSKEEVPYTHLAKFP
ENAEDELVOETQHLFLOYKQKIDDKIYCPASVILASVAQAKGYDPSVHR
GFLAOEELPKRVINLQMPMEKEITANYAHRBRADAEMLTKAOLEMWG
VNFPIRNKKGTLELDALGLRIYDENLTKISPPNMEINISDSDEFIKPL

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-364J6"
/clone_lib="RPc1-11"
859. .982
/repeat_region
/rpt_family="MER94"
1101. .1234
/rpt_family="Alu"
1235. .1541
/rpt_family="Alu"
1342. .1674
/rpt_family="Alu"
1675. .1817
/rpt_family="(GAGAA)n"
2423. .2856
/misc_feature
/note="match to EST AA457626 (NID:g2180346) aa89610.r1"
repeat_region
3122. .3308
/rpt_family="MER1_type"
4452. .4500
/rpt_family="GA-rich"
4626. .5212
/rpt_family="L1"
5583. .5663
/rpt_family="L2"
5683. .5804
/rpt_family="L2"
5839. .6177
/rpt_family="L2"
8592. .8703
/rpt_family="MALR"
8770. .8994
/rpt_family="MALR"
9506. .9681
/rpt_family="MIR"
9682. .9973
/rpt_family="(CAT)n"
9990. .10172
/rpt_family="(CAT)n"
10210. .10389
/rpt_family="(CAT)n"
10404. .10580

```

```

repeat_region      /rpt_family="(CAT)n"
10598..11147
/rpt_family="(CAT)n"
repeat_region      /rpt_family="(CAT)n"
11159..11338
/rpt_family="(CAT)n"
repeat_region      /rpt_family="(CAT)n"
11350..11572
/rpt_family="(CAT)n"
repeat_region      /rpt_family="(CAT)n"
11514..11954
/rpt_family="(CAT)n"
misc-feature       /rpt_family="(CAT)n"
11558..12135
/note="similar to EST AA425721 (NTD:92106441) zw49b10.s1"
misc-feature       /note="similar to EST AA425721 (NTD:92106441) zw49b10.s1"
13250..13597
/note="match to EST AA427791 (NTD:92111607) zw49b10.r1"
repeat_region      /rpt_family="(L2)"
13735..14214
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
15036..15174
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
15560..15686
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
15777..15960
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
15981..16065
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
16066..16366
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
16367..16539
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
17008..17175
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
17176..17463
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
17464..18111
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
19413..19447
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
19766..19856
/rpt_family="(L2)"
misc-feature       /note="match to EST AA494044 (NTD:92233885) ng60e10.s1"
21086..21424
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
21133..21172
/rpt_family="(L2)"
misc-feature       /note="match to EST AA494044 (NTD:92233885) ng60e10.s1"
21430..21451
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
21595..21634
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
21911..21954
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
24210..24476
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
26241..26264
/rpt_family="(L2)"
misc-feature       /note="match to EST AA476160 (NTD:94329205) tn62h08.x1"
26751..26970
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
28700..28725
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
29474..29763
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
30021..30386
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
31035..31121
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
31494..31719
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
31720..31758
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
32071..32178
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
32804..32511
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
33334..33375
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
33422..33441
/rpt_family="(L2)"

```

```

repeat_region      33679..33739
/rpt_family="(L2)"
repeat_region      34762..34820
/rpt_family="(L2)"
Query Match      1.5%; Score 20; DB 9; Length 156374;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1181 gagctgcagaattcaagttg 1200
|||||
Db 52151 GAGCTGCAGATTCAAGTTG 52132
|||||

```

```

RESULT 35
AL596283
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Project Information
Center project name: h0257N2
----- Summary Statistics -----
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 158737 bases at least Q40
Consensus quality: 159906 bases at least Q30
Consensus quality: 160628 bases at least Q20
Insert size: 161323; sum-of-contrigs
Insert size: 165641; 3.6% error; agarose-tp
Quality coverage: 6.59x in Q20 bases; sum-of-contrigs Quality
coverage: 6.45x in Q20 bases; agarose-tp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

FEATURES
SOURCE
Location/Qualifiers
1..163323
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="XXbac-257N2"
/clone_lib="NOD mouse library"
1..6428
/note="assembly-fragment:01729"
fragment_chain:1
6529..14784
/note="assembly-fragment:02816"
fragment_chain:1
14885..37995
/note="assembly-fragment:01187.0"
38096..55258
/note="assembly-fragment:00197"
fragment_chain:2

```

```

misc_feature      52629..71295
                  /note="assembly_fragment:01892
                  fragment_chain:2"
misc_feature      71396..78245
                  /note="assembly_fragment:02371
                  fragment_chain:2"
misc_feature      78346..103280
                  /note="assembly_fragment:01370
                  fragment_chain:2"
misc_feature      103381..135684
                  /note="assembly_fragment:01702
                  fragment_chain:2"
misc_feature      135785..146174
                  /note="assembly_fragment:02231
                  fragment_chain:2"
misc_feature      146275..160087
                  /note="assembly_fragment:03006
                  fragment_chain:2"
misc_feature      160188..162323
                  /note="assembly_fragment:01168
                  fragment_chain:2
                  clone_end:sp6
                  vector_side:right"
BASE COUNT      45002 a 33182 c 33907 g 49227 t 1005 others
ORIGIN
Query Match      1.5%; Score 20; DB 2; Length 162323;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      329 gtgacattatcacagactcta 348
      |||||||
Db 128540 GTGACATTATCACAGACTCTA 128559

RESULT 36
AC073556/c      163189 bp      DNA      16-AUG-2001
LOCUS
DEFINITION      Oryza sativa chromosome 3 clone OSJNBa0091P11. *** SEQUENCING IN
PROGRESS ***. 1 ordered pieces.
ACCESSION      AC073556
VERSION      AC073556.8 GI:15193445
KEYWORDS      HTG; HTGS_PHASE2.
SOURCE      Oryza sativa.
ORGANISM      Oryza sativa
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
               Ehrhartoideae; Oryzoideae; Oryza.
REFERENCE      1 (bases 1 to 163189)
               Buell, R., Hsiao, J., Zismann, V., Moffat, K.M., Hill, J.,
               Gansberger, K., Burgess, S., Jarrahi, B., Shwartsbeyn, M., Brenner, M.,
               Ciescio, A., Pal, G., Vanaken, S., Hansen, C., Utecherbach, T.,
               Feldblyum, T., Khalak, H.G., Yuan, Q., Quackenbush, J., White, O.,
               Salzberg, S. and Fraser, C.
               Oryza sativa ssp. japonica cv. Nipponbare OSJNBa0091P11 BAC genomic
               sequence
               Unpublished
               2 (bases 1 to 163189)
               Buell, R.
               Direct Submission
               Submitted (23-JUN-2000) The Institute for Genomic Research, 9712
               Medical Center Dr., Rockville, MD 20850, USA
               On Aug 16, 2001 this sequence version replaced gi:14993751.
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 1 contigs. Gaps between the contigs
               * are represented as runs of N. The order of the pieces
               * is believed to be correct as given, however the sizes
               * of the gaps between them are based on estimates that have
               * provided by the submitter.
               * This sequence will be replaced
               * by the finished sequence as soon as it is available and
               * the accession number will be preserved.

```

```

FEATURES
Source
      * 1 163189: contig of 163189 bp in length.
      Location/Qualifiers
          1..163189
             /organism="Oryza sativa"
             /cultivar="Nipponbare"
             /sub_species="japonica"
             /db_xref="taxon:4530"
             /chromosome="3"
             /clone="OSJNBa0091P11"
BASE COUNT      47068 a 35531 c 35052 g 45534 t 4 others
ORIGIN
Query Match      1.5%; Score 20; DB 2; Length 163189;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      905 gtgctatttctttaaagt 924
      |||||||
Db 73824 GTGCTATTCTTTTAAAGT 73805

```

```

RESULT 37
AC016845/c      170294 bp      DNA      11-AUG-2001
LOCUS
DEFINITION      Homo sapiens clone RP11-5E23. *** SEQUENCING IN PROGRESS ***. 5
unordered pieces.
ACCESSION      AC016845
VERSION      AC016845.6 GI:15148233
KEYWORDS      HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 170294)
               Birren, B., Linton, L., Nusbaum, C. and Lander, E.
               Unpublished
               2 (bases 1 to 170294)
               Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
               Balding, J., Barina, N., Beckert, R., Boguslavsky, L., Boukhalter, B.,
               Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
               Cooke, P., DeRellano, K., Dewar, K., Domingo, M., Donahue, L., Doyle, M.,
               Ferreira, P., Fitzhugh, W., Forrest, C., Funke, K., Gage, D.,
               Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
               Howland, J., Johnson, R., Jones, C., Kane, L., Karas, A., Klein, J.,
               Lechoczy, J., Liu, C., Locke, K., Macdonald, P., Marquis, N.,
               McEwan, P., McGurk, A., McKernan, K., McDonald, J., Meldrum, J.,
               Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
               Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
               Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
               Tesigye, S., Tittell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
               Wyman, D., Ye, W., J., Zimmer, A. and Zody, M.
               Direct Submission
               Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
               On Aug 11, 2001 this sequence version replaced gi:11990706.
               All repeats were identified using RepeatMasker:
               http://ftp.genome.washington.edu/RV/RepeatMasker.html
               ----- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: M18R
               Web site: http://www.seg.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ----- Project Information
               Center project name: 5_E_23
               Center clone name: 5_E_23
               -----
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 5 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as

```



```

STS
repeat_region /rpt_family="L1" complement(5091..5707)
                /db_xref="GI:6126790"
5329..5549
repeat_region /rpt_family="L1"
5578..6135
repeat_region /rpt_family="L1"
6230..6913
repeat_region /rpt_family="L1"
6914..7288
repeat_region /rpt_family="Retroviral"
7365..7613
repeat_region /rpt_family="L1"
7614..7916
repeat_region /rpt_family="Alu"
7917..7948
repeat_region /rpt_family="L1"
7949..8246
repeat_region /rpt_family="Alu"
8247..8510
repeat_region /rpt_family="L1"
8640..8933
repeat_region /rpt_family="Alu"
8937..9246
repeat_region /rpt_family="Alu"
9247..9276
repeat_region /rpt_family="(TAAA)n"
9567..10019
repeat_region /rpt_family="MALR"
10220..10387
repeat_region /rpt_family="Alu"
10388..10433
repeat_region /rpt_family="(TTTA)n"
10434..10713
repeat_region /rpt_family="Alu"
10714..10840
repeat_region /rpt_family="Alu"
10843..11060
repeat_region /rpt_family="L1"
11325..11357
repeat_region /rpt_family="(TTTA)n"
11358..11641
repeat_region /rpt_family="Alu"
11823..11916
repeat_region /rpt_family="L1"
11917..11937
repeat_region /rpt_family="(TTTC)n"
11938..12224
repeat_region /rpt_family="Alu"
12225..12361
repeat_region /rpt_family="L1"
12394..12526
repeat_region /rpt_family="Alu"
12527..12816
repeat_region /rpt_family="Alu"
12817..12933
repeat_region /rpt_family="Alu"
13066..13088
repeat_region /rpt_family="(TTTC)n"
13089..13237
repeat_region /rpt_family="Alu"
13402..13427
repeat_region /rpt_family="(T)n"
13433..13681
repeat_region /rpt_family="Alu"
13841..14021
repeat_region /rpt_family="MERL_type"
14045..14178
repeat_region /rpt_family="L1"
14221..14269
repeat_region /rpt_family="Alu"
14337..14616
repeat_region /rpt_family="Alu"

Query Match 1.5% Score 20; DB 9; Length 173893;
Best Local Similarity 100.0%; Pred No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 784 ttggaagattcacttacc 803
Db 45927 TTGGAAGATTCACTTACCA 45946

RESULT 39
AC022812
LOCUS AC022812 174596 bp DNA HTG 12-MAR-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-745f13 map 6, WORKING DRAFT
SEQUENCE, 24 unordered pieces.
ACCESSION AC022812
VERSION AC022812.2 GI:7229851
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 174596)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 6, clone RP11-745f13
2 (bases 1 to 174596)
Unpublished
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Bede,F.,
Boguski,K., Bouckhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRubeis,K., Dewar,K., Domino,M., Doyle,M., Fenebor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heathcote,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGuck,A., McKernan,K.,
McPheters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,J.,
Stojanovic,N., Subramanian,A., Talamas,J., Teisaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6922250.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6096

```

Center clone name: 745_F_13
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 161291 bases at least Q40
 Consensus quality: 168261 bases at least Q30
 Consensus quality: 170761 bases at least Q20
 Insert size: 182000; agarose-fp
 Insert size: 172296; sum-of-contrigs
 Quality coverage: 3.8 in Q20 bases; agarose-fp
 Quality coverage: 4.0 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 24 contrigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contrigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1284: contrig of 1284 bp in length
 1285 1384: gap of 100 bp
 1385 2706: contrig of 1322 bp in length
 2707 2806: gap of 100 bp
 2807 4528: contrig of 1722 bp in length
 4529 4628: gap of 100 bp
 4629 6469: contrig of 1841 bp in length
 6470 6569: gap of 100 bp
 6570 10731: contrig of 4162 bp in length
 10732 10831: gap of 100 bp
 10832 13405: contrig of 2574 bp in length
 13406 13505: gap of 100 bp
 13506 16970: contrig of 3465 bp in length
 16971 17070: gap of 100 bp
 17071 20522: contrig of 3452 bp in length
 20523 20622: gap of 100 bp
 20623 23195: contrig of 2573 bp in length
 23196 23295: gap of 100 bp
 23296 27465: contrig of 4170 bp in length
 27466 27565: gap of 100 bp
 27566 32567: contrig of 5002 bp in length
 32568 32667: gap of 100 bp
 32668 37897: contrig of 5230 bp in length
 37898 37997: gap of 100 bp
 37998 45060: contrig of 7063 bp in length
 45061 45160: gap of 100 bp
 45161 50594: contrig of 5433 bp in length
 50595 50694: gap of 100 bp
 50695 58198: contrig of 7504 bp in length
 58199 58298: gap of 100 bp
 58299 68303: contrig of 10005 bp in length
 68304 68403: gap of 100 bp
 68404 77724: contrig of 9321 bp in length
 77725 77824: gap of 100 bp
 77825 86591: contrig of 8767 bp in length
 86592 86691: gap of 100 bp
 86692 95521: contrig of 8830 bp in length
 95522 95621: gap of 100 bp
 95622 105952: contrig of 10331 bp in length
 105953 106052: gap of 100 bp
 106053 117439: contrig of 11387 bp in length
 117440 117539: gap of 100 bp
 117540 132556: contrig of 15017 bp in length
 132557 132656: gap of 100 bp
 132657 148474: contrig of 15818 bp in length
 148475 148574: gap of 100 bp
 148575 174596: contrig of 26022 bp in length.

FEATURES
 SOURCE
 1. 174596
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"

/map="6"
 /clone="RP11-745F13"
 /clone_1b="RP11 Human Male BAC"
 1. 1284
 /note="assembly-fragment"
 1385 2706
 /note="assembly-fragment"
 2807 4528
 /note="assembly-fragment"
 4629 6469
 /note="assembly-fragment"
 6570 10731
 /note="assembly-fragment"
 10832 13405
 /note="assembly-fragment"
 13506 16970
 /note="assembly-fragment"
 17071 20522
 /note="assembly-fragment"
 20623 23195
 /note="assembly-fragment"
 /note="assembly-fragment"
 clone_end:T7
 vector_side:left
 23296 27465
 /note="assembly-fragment"
 27566 32567
 /note="assembly-fragment"
 32668 37897
 /note="assembly-fragment"
 37998 45060
 /note="assembly-fragment"
 45161 50594
 /note="assembly-fragment"
 clone_end:SP6
 vector_side:left
 50695 58198
 /note="assembly-fragment"
 58299 68303
 /note="assembly-fragment"
 68404 77724
 /note="assembly-fragment"
 77825 86591
 /note="assembly-fragment"
 86692 95521
 /note="assembly-fragment"
 95622 105952
 /note="assembly-fragment"
 106053 117439
 /note="assembly-fragment"
 117540 132556
 /note="assembly-fragment"
 132657 148474
 /note="assembly-fragment"
 148575 174596
 /note="assembly-fragment"

BASE COUNT 54131 a 33318 c 33201 g 51641 t 2305 others
 ORIGIN

Query Match 1.5%; Score 20; DB 2; Length 174596;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1226 ctaataattgattcctaag 1245
 |||||
 Db 57509 CTAATAATTGATTCTAAG 57528

RESULT 40
 AC007486 174725 bp DNA PRI 05-MAY-1999
 LOCUS AC007486/c
 DEFINITION Homo sapiens clone D1015P16A, complete sequence.
 ACCESSION AC007486

```

VERSION      AC007486.1  GI:4753270
KEYWORDS
SOURCE       human
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE       1 (bases 1 to 174725)
JOURNAL     Waterston, R.H.
REFERENCE    The sequence of Homo sapiens clone
AUTHORS     Unpublished
TITLE       2 (bases 1 to 174725)
JOURNAL     Waterston, R.H.
REFERENCE    Submitted (05-MAY-1999) Genome Sequencing Center, Washington
AUTHORS     University School of Medicine, 4444 Forest Park Parkway, St. Louis,
TITLE       MO 63108, USA
JOURNAL
FEATURES
Source       1..174725
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="DJ1015P16A"
BASE COUNT   58080 a 34168 c 31762 g 50715 t
ORIGIN
Query Match      1.5%; Score 20; DB 9; Length 174725;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 tattcaatgtattttaa 27
|||||
Db 162503 TATTCAAATGTTATTTTAA 162484

RESULT 41
AC013261/C 175092 bp DNA HTG 30-OCT-2000
LOCUS Homo sapiens clone RP11-11E18, WORKING DRAFT SEQUENCE, 2 ordered
DEFINITION
Pieces
AC013261
AC013261.3 GI:11038541
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 175092)
JOURNAL Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE Homo sapiens, clone RP11-11E18
AUTHORS Unpublished
TITLE 2 (bases 1 to 175092)
JOURNAL Unpublished
AUTHORS
Bairren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Bouckgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Dearrellan, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Headford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,
Lehoczky, J., Lien, C., Locke, K., MacDonald, P., Marquis, N.,
McKernan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testafaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (05-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 30, 2000 this sequence version replaced g1:7107901.
All repeats were identified using RepeatMasker:
Smit, A.F., A. Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: M13R
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3142
Center clone name: 11_E-18
----- Summary Statistics
Sequencing vector: M13; M77815; 48% of reads
Sequencing vector: plasmid; n/a; 52% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 174862 bases at least Q40
Consensus quality: 174949 bases at least Q20
Consensus quality: 174968 bases at least Q20
Insert size: 176000; agarose-tp
Insert size: 174992; sum-of-contigs
Quality coverage: 10.9 in Q20 bases; agarose-tp
Quality coverage: 11.0 in Q20
NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 86418: contig of 86418 bp in length
* 86419 86518: gap of 100 bp
* 86519 175092: contig of 88574 bp in length.
Location/Qualifiers
1..175092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-11E18"
/clone_lib="RP11-11 Human Male BAC"
1..86418
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
86519..175092
/note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 50202 a 34212 c 35769 g 54809 t 100 others
ORIGIN
Query Match      1.5%; Score 20; DB 2; Length 175092;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1226 ctaataatgtattcraag 1245
|||||
Db 39221 CTAAATAATGATTCCTAAG 39202

RESULT 42
AC090822 181994 bp DNA HTG 27-MAY-2001
LOCUS Homo sapiens chromosome 11 clone CTD-2341A5 map 11, WORKING DRAFT
DEFINITION
SEQUENCE 6 unordered pieces.
AC090822
AC090822.2 GI:14210584
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 181994)
JOURNAL Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE Homo sapiens chromosome 11, clone CTD-2341A5

```

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 181994)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,
Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Fero, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J.,
Gardyna, S., Glende, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., Larocque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., Maclean, C., MacDonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Meldrum, J., Menus, L., Mihova, T., Mlenga, Y.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retter, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P.,
Souqnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Travers, M., Travis, N., Trifilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zemek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Direct Submission
Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 27, 2001 this sequence version replaced g1:13270691.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L12835

Center clone name: 2341.A.5

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179026 bases at least Q40
Consensus quality: 179649 bases at least Q30
Consensus quality: 179945 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 181494; sum-of-contigs
Quality coverage: 10.9 in Q20 bases; agarose-fp
Quality coverage: 10.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 609: contig of 609 bp in length
* 610 709: gap of 100 bp
* 710 1362: contig of 653 bp in length
* 1363 1462: gap of 100 bp
* 1463 8654: contig of 7192 bp in length
* 8655 8754: gap of 100 bp
* 8755 24527: contig of 15773 bp in length
* 24528 24627: gap of 100 bp
* 24628 82493: contig of 57866 bp in length
* 82494 82594: gap of 100 bp
* 82594 181994: contig of 99401 bp in length.

FEATURES
Source

Location/Qualifiers
1..181994
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="11"
/map="11"
/clone="CPD-2341A5"
/clone_lib="CITD Human BAC"
1..609
/note="assembly_fragment"
710..1362
/note="assembly_fragment"
1463..8654
/note="assembly_fragment"
8755..24527
/note="assembly_fragment"
24628..82493
/note="assembly_fragment"
82594..181994
/note="assembly_fragment"
BASE COUNT 56084 a 33254 c 33265 g 58889 t 502 others
ORIGIN

Query Match 1.5%; Score 20; DB 2; Length 181994;
Best Local Similarity 100.0%; Pired. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 caaatgtattttaacata 31
DB 70914 CAAATGTTATTTTAAACATA 70933

RESULT 43
AC026544/C
LOCUS
DEFINITION Homo sapiens chromosome 6 clone RP11-163122 map 6, WORKING DRAFT
ACCESSION AC026544
VERSION AC026544.2 GI:7528136
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 182807)
Birken, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 6, clone RP11-163122
Unpublished
2 (bases 1 to 182807)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burnett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glende, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, A., Larocque, K., Lamazares, R., Landers, T., Lehoczky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenga, Y., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Trifilio, A., Travers, M., Trifilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 8, 2000 this sequence version replaced g1:7283252.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 17232
Center clone name: 163.I.22
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16531 bases at least Q40
Consensus quality: 17394 bases at least Q30
Consensus quality: 17715 bases at least Q20
Insert size: 18800; agarose-fp
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1088: contig of 1088 bp in length
* 1089 1188: gap of 100 bp
* 1189 2428: contig of 1240 bp in length
* 2429 2528: gap of 100 bp
* 2529 3878: contig of 1350 bp in length
* 3879 3978: gap of 100 bp
* 3979 5268: contig of 1290 bp in length
* 5269 5368: gap of 100 bp
* 5369 7753: contig of 2385 bp in length
* 7754 7853: gap of 100 bp
* 7854 9680: contig of 1837 bp in length
* 9691 9790: gap of 100 bp
* 9791 11570: contig of 1780 bp in length
* 11571 11670: gap of 100 bp
* 11671 13719: contig of 2049 bp in length
* 13720 13819: gap of 100 bp
* 13820 16027: contig of 2208 bp in length
* 16028 16127: gap of 100 bp
* 16128 18113: contig of 1986 bp in length
* 18114 18213: gap of 100 bp
* 18214 19867: contig of 1654 bp in length
* 19868 19967: gap of 100 bp
* 19968 21902: contig of 1935 bp in length
* 21903 22002: gap of 100 bp
* 22003 24557: contig of 2555 bp in length
* 24558 24657: gap of 100 bp
* 24658 27245: contig of 2588 bp in length
* 27246 27345: gap of 100 bp
* 27346 30656: contig of 3311 bp in length
* 30657 30756: gap of 100 bp
* 30757 33365: contig of 2609 bp in length
* 33366 33465: gap of 100 bp
* 33466 35816: contig of 2351 bp in length
* 35817 35916: gap of 100 bp
* 35917 38408: contig of 2492 bp in length
* 38409 38508: gap of 100 bp
* 38509 42615: contig of 4107 bp in length
* 42616 42715: gap of 100 bp
* 42716 47091: contig of 4375 bp in length
* 47092 47191: gap of 100 bp
* 47192 53079: contig of 5888 bp in length
* 53080 53179: gap of 100 bp
* 53180 56645: contig of 3466 bp in length
* 56646 56745: gap of 100 bp

```

FEATURES

Source

```

* 56746 63020: contig of 6275 bp in length
* 63021 63120: gap of 100 bp
* 63121 71045: contig of 7925 bp in length
* 71046 71145: gap of 100 bp
* 71146 76850: contig of 5705 bp in length
* 76851 76950: gap of 100 bp
* 76951 85447: contig of 8497 bp in length
* 85448 85547: gap of 100 bp
* 85548 91907: contig of 6360 bp in length
* 91908 92007: gap of 100 bp
* 92008 98725: contig of 6718 bp in length
* 98726 98825: gap of 100 bp
* 98826 107166: contig of 8341 bp in length
* 107167 107266: gap of 100 bp
* 107267 115271: contig of 8005 bp in length
* 115272 115371: gap of 100 bp
* 115372 123500: contig of 8129 bp in length
* 123501 123600: gap of 100 bp
* 123601 133338: contig of 9738 bp in length
* 133339 133438: gap of 100 bp
* 133439 142253: contig of 8815 bp in length
* 142254 142353: gap of 100 bp
* 142354 155872: contig of 13519 bp in length
* 155873 155972: gap of 100 bp
* 155973 167753: contig of 11761 bp in length
* 167754 167853: gap of 100 bp
* 167854 182807: contig of 14954 bp in length.
*
* Location/Qualifiers
* 1..182807
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /chromosome="6"
*   /map="6"
*   /clone="RP11-163122"
*   /clone_jib="RPCT-11 Human Male BAC"
*   /note="assembly-fragment"
* 1189..2428
*   /note="assembly-fragment"
* 2529..3878
*   /note="assembly-fragment"
* 3979..5268
*   /note="assembly-fragment"
*   /note="assembly-fragment"
*   /vector_side="right"
* 5369..7753
*   /note="assembly-fragment"
* 7854..9690
*   /note="assembly-fragment"
* 9791..11570
*   /note="assembly-fragment"
* 11671..13719
*   /note="assembly-fragment"
* 13820..16027
*   /note="assembly-fragment"
* 16128..18113
*   /note="assembly-fragment"
* 18214..19867
*   /note="assembly-fragment"
* 19968..21902
*   /note="assembly-fragment"
* 22003..24557
*   /note="assembly-fragment"
* 24658..27245
*   /note="assembly-fragment"
* 27346..30656
*   /note="assembly-fragment"
* 30757..33365
*   /note="assembly-fragment"
* 33466..35816
*   /note="assembly-fragment"
* 35917..38408
*   /note="assembly-fragment"

```

```

misc_feature      38509..42615
                  /note="assembly-fragment"
misc_feature      42716..47091
                  /note="assembly-fragment"
misc_feature      47192..53079
                  /note="assembly-fragment"
misc_feature      53180..56645
                  /note="assembly-fragment"
misc_feature      56746..63020
                  /note="assembly-fragment"
misc_feature      63321..71045
                  /note="assembly-fragment"

Query Match      1.5%; Score 20; DB 2; Length 182807;
Best Local Similarity 100.0%; Prid. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      119      aaaaagacattacttggtt 138
Db 153217      AAAAAGACATTACTTGSTT 153198
|||||
|||||

RESULT  44
AC080188      183147 bp      DNA      HTG      26-OCT-2000
LOCUS      Homo sapiens chromosome 4 clone RPL1-635FL1, WORKING DRAFT SEQUENCE,
DEFINITION      40 unordered pieces.
ACCESSION      AC080188
VERSION      AC080188.4 GI:11024950
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
              Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 183147)
              Waterston, R.H.
              The sequence of Homo sapiens clone
              Unpublished
              2 (bases 1 to 183147)
              Waterston, R.H.
              Direct Submission
              Submitted (28-SEP-2000) Genome Sequencing Center, Washington
              University School of Medicine, 444 Forest Park Parkway, St. Louis,
              MO 63108 USA
              On Oct 26, 2000 this sequence version replaced gi:10945799.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center Project name: H.NH0635L01
----- Summary Statistics -----
Sequencing vector: MJ3, 1008
Sequencing vector: Plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158849 bases at least Q40
Consensus quality: 167365 bases at least Q30
Consensus quality: 171079 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 178294; sum-of-contrigs
Quality coverage: 3.58 in Q20 bases; agarose-fp
Quality coverage: 3.81 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

*	*	be preserved.	
*	1	1171: contig of 1171 bp in length	
*	1172	1271: gap of unknown length	
*	1372	2638: contig of 1367 bp in length	
*	2639	2738: gap of unknown length	
*	2739	3913: contig of 1175 bp in length	
*	3914	4013: gap of unknown length	
*	4014	6045: contig of 2032 bp in length	
*	6046	6145: gap of unknown length	
*	6146	7912: contig of 1767 bp in length	
*	7913	8012: gap of unknown length	
*	8013	9955: contig of 1943 bp in length	
*	9956	10055: gap of unknown length	
*	10056	12248: contig of 2193 bp in length	
*	12249	12348: gap of unknown length	
*	12349	12175: contig of 2827 bp in length	
*	15176	15275: gap of unknown length	
*	15276	14783: contig of 2208 bp in length	
*	17464	17583: gap of unknown length	
*	17584	19758: contig of 2175 bp in length	
*	19759	19858: gap of unknown length	
*	19859	21922: contig of 2064 bp in length	
*	21923	22022: gap of unknown length	
*	22023	23879: contig of 1837 bp in length	
*	23880	23979: gap of unknown length	
*	23980	26337: contig of 2458 bp in length	
*	26337	26537: gap of unknown length	
*	26438	28826: contig of 2289 bp in length	
*	26538	28926: gap of unknown length	
*	28827	31019: contig of 2093 bp in length	
*	28927	31119: gap of unknown length	
*	31020	33119: gap of unknown length	
*	31120	33924: contig of 2805 bp in length	
*	33925	34024: gap of unknown length	
*	34025	37493: contig of 3469 bp in length	
*	37494	37593: gap of unknown length	
*	37594	40607: contig of 3014 bp in length	
*	40608	40707: gap of unknown length	
*	40708	43763: contig of 3056 bp in length	
*	43764	43863: gap of unknown length	
*	43864	48779: contig of 4916 bp in length	
*	48780	48879: gap of unknown length	
*	48880	53410: contig of 4531 bp in length	
*	53411	53510: gap of unknown length	
*	53511	57610: contig of 4100 bp in length	
*	57610	57710: gap of unknown length	
*	57710	58653: contig of 953 bp in length	
*	58661	58763: gap of unknown length	
*	58764	63998: contig of 5135 bp in length	
*	63999	65998: gap of unknown length	
*	65999	70378: contig of 6380 bp in length	
*	70379	70478: gap of unknown length	
*	70479	76232: contig of 5774 bp in length	
*	76233	76352: gap of unknown length	
*	76353	82864: contig of 6512 bp in length	
*	82865	82964: gap of unknown length	
*	82965	89557: contig of 6633 bp in length	
*	85598	89557: gap of unknown length	
*	85698	95761: contig of 6064 bp in length	
*	85762	95861: gap of unknown length	
*	95862	101602: contig of 5741 bp in length	
*	101603	101702: gap of unknown length	
*	101703	108432: contig of 6720 bp in length	
*	108433	108532: gap of unknown length	
*	108523	113599: contig of 5077 bp in length	
*	113599	113699: gap of unknown length	
*	113700	121769: contig of 8095 bp in length	
*	121795	121894: gap of unknown length	
*	121895	130939: contig of 9045 bp in length	
*	130940	131039: gap of unknown length	
*	131040	138924: contig of 7885 bp in length	
*	139925	139024: gap of unknown length	
*	139925	147837: contig of 8813 bp in length	
*	147838	147937: gap of unknown length	

```

* 147938 158381: contig of 10444 bp in length
* 158382 158481: gap of unknown length
* 158482 167102: contig of 8621 bp in length
* 167103 167202: gap of unknown length
* 167203 181613: contig of 14411 bp in length
* 181614 181713: gap of unknown length
* 181714 183147: contig of 1434 bp in length.
      location/qualifiers
1. 183147
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="4"
   /clone="RP11-635H1"
1. 1171
misc_feature /note="assembly_name:Contig13"
1272. 2638
misc_feature /note="assembly_name:Contig14"
2739. 3913
misc_feature /note="assembly_name:Contig16"
4014. 6045
misc_feature /note="assembly_name:Contig17"
6146. 7912
misc_feature /note="assembly_name:Contig18"
8013. 9955
misc_feature /note="assembly_name:Contig19"
10056. 12248
misc_feature /note="assembly_name:Contig20"
12349. 15175
misc_feature /note="assembly_name:Contig21"
15276. 17483
misc_feature /note="assembly_name:Contig22"
17594. 19758
misc_feature /note="assembly_name:Contig23"
19859. 21922
misc_feature /note="assembly_name:Contig24"
22023. 23879
misc_feature /note="assembly_name:Contig25"
23960. 26437
misc_feature /note="assembly_name:Contig26"
26538. 28826
misc_feature /note="assembly_name:Contig27"
28927. 31019
misc_feature /note="assembly_name:Contig28"
31120. 33924
misc_feature /note="assembly_name:Contig29"
34025. 37493
misc_feature /note="assembly_name:Contig30"
37594. 40607
misc_feature /note="assembly_name:Contig31"
40708. 43763
misc_feature /note="assembly_name:Contig32"
43864. 48779
misc_feature /note="assembly_name:Contig33"
48880. 53410
misc_feature /note="assembly_name:Contig34"
53511. 57610
misc_feature /note="assembly_name:Contig35
clone_end:sp6
vector_side:right"
57711. 58663
misc_feature /note="assembly_name:Contig12"
58764. 63898
misc_feature /note="assembly_name:Contig36"
63999. 70378
misc_feature /note="assembly_name:Contig37"
70479. 76252
misc_feature /note="assembly_name:Contig38"
76353. 82864
misc_feature /note="assembly_name:Contig39"
82965. 89597
misc_feature /note="assembly_name:Contig40"
89698. 95761
misc_feature /note="assembly_name:Contig41"

```

```

misc_feature 95862. 101602 /note="assembly_name:Contig42"
misc_feature 101703. 108423 /note="assembly_name:Contig43"
misc_feature 108523. 113599

Query Match 1.5%; Score 20; DB 2; Length 183147;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 aaaaagacattacttggt 138
Db 100133 AAAAGACATTACTTGTT 100152

RESULT 45
AC079856/c 185421 bp DNA HTG 02-FEB-2001
LOCUS Homo sapiens chromosome RPCL-11 clone RP11-521B4, WORKING DRAFT
DEFINITION SEQUENCE, 46 unordered pieces.
ACCESSION AC079856
VERSION AC079856.4 GI:12656844
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 185421)
REFERENCE 1
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL unpublished
2 (bases 1 to 185421)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 2, 2001 this sequence version replaced gi:11560300.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0521B04
----- Summary Statistics -----
Sequencing vector: MJ3; 100%
Chemistry: Dye-Primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 162590 bases at least Q40
Consensus quality: 169535 bases at least Q30
Consensus quality: 172901 bases at least Q20
Insert size: 181000; agarose-fp
Insert size: 180921; sum-of-ctnigs
Quality coverage: 3.10 in Q20 bases; sum-of-ctnigs
Quality coverage: 3.51 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1522: contig of 1522 bp in length
* 1523 1622: gap of unknown length
* 1623 2684: contig of 1062 bp in length
* 2685 2784: gap of unknown length
* 2785 4631: contig of 1847 bp in length
* 4632 4731: gap of unknown length

```

```

* 4732 6051: contig of 1320 bp in length
* 6052 6151: gap of unknown length
* 6152 7586: contig of 1435 bp in length
* 7587 7687 10059: gap of unknown length
* 7687 10059: contig of 2373 bp in length
* 10060 10159: gap of unknown length
* 10160 12215: contig of 2056 bp in length
* 12216 12315: gap of unknown length
* 12316 14002: contig of 1687 bp in length
* 14003 14102: gap of unknown length
* 14103 15350: contig of 1248 bp in length
* 15351 15450: gap of unknown length
* 15451 16663: contig of 1213 bp in length
* 16664 16763: gap of unknown length
* 16764 19156: contig of 2393 bp in length
* 19157 19256: gap of unknown length
* 19257 21199: contig of 1943 bp in length
* 21200 21299: gap of unknown length
* 21300 23500: contig of 2201 bp in length
* 23501 23600: gap of unknown length
* 23601 25776: contig of 2176 bp in length
* 25777 25876: gap of unknown length
* 25877 27788: contig of 1912 bp in length
* 27789 27888: gap of unknown length
* 27889 30449: contig of 2561 bp in length
* 30450 30549: gap of unknown length
* 30550 33166: contig of 2617 bp in length
* 33167 33266: gap of unknown length
* 33267 36506: contig of 3240 bp in length
* 36507 36606: gap of unknown length
* 36607 39573: contig of 2967 bp in length
* 39574 39674: gap of unknown length
* 39674 42099: contig of 2426 bp in length
* 42100 42199: gap of unknown length
* 42200 45279: contig of 3080 bp in length
* 45280 45379: gap of unknown length
* 45380 48433: contig of 3054 bp in length
* 48434 48533: gap of unknown length
* 48534 51920: contig of 3387 bp in length
* 51921 52020: gap of unknown length
* 52021 55180: contig of 3160 bp in length
* 55181 55280: gap of unknown length
* 55281 57733: contig of 2433 bp in length
* 57734 60953: gap of unknown length
* 60954 61053: gap of unknown length
* 61054 64482: contig of 3429 bp in length
* 64483 64582: gap of unknown length
* 64583 67979: contig of 3397 bp in length
* 67980 68079: gap of unknown length
* 68080 71233: contig of 3154 bp in length
* 71234 71333: gap of unknown length
* 71334 75409: contig of 4076 bp in length
* 75410 75509: gap of unknown length
* 75510 80821: contig of 5312 bp in length
* 80822 80921: gap of unknown length
* 80922 85178: contig of 4257 bp in length
* 85179 85278: gap of unknown length
* 85279 89505: contig of 4227 bp in length
* 89506 89605: gap of unknown length
* 89606 94860: contig of 5255 bp in length
* 94861 94960: gap of unknown length
* 94961 100923: contig of 5963 bp in length
* 100924 101023: gap of unknown length
* 101024 107351: contig of 6328 bp in length
* 107352 107451: gap of unknown length
* 107452 113862: contig of 6411 bp in length
* 113863 113962: gap of unknown length
* 113963 118976: contig of 5014 bp in length
* 118977 119076: gap of unknown length
* 119077 124602: contig of 5526 bp in length
* 124603 124702: gap of unknown length
* 124703 131096: contig of 6394 bp in length

```

FEATURES

source

```

* 131097 131196: gap of unknown length
* 131197 138541: contig of 7345 bp in length
* 138542 138641: gap of unknown length
* 138642 146000: contig of 7359 bp in length
* 146001 146100: gap of unknown length
* 146101 154773: contig of 8673 bp in length
* 154774 154873: gap of unknown length
* 154874 163845: contig of 8972 bp in length
* 163846 163945: gap of unknown length
* 163946 175150: contig of 11205 bp in length
* 175151 175250: gap of unknown length
* 175251 185421: contig of 10171 bp in length.

FEATURES
source
1. 185421 "Homo sapiens"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="PC1-11"
/clone="RP11-521B4"
1. 1522
/note="assembly_name:Contig13"
1623. 2684
/note="assembly_name:Contig14"
2785. 4631
/note="assembly_name:Contig17"
4732. 6051
/note="assembly_name:Contig18"
6152. 7586
/note="assembly_name:Contig19"
7687. 10059
/note="assembly_name:Contig20"
10160. 12215
/note="assembly_name:Contig21"
12316. 14002
/note="assembly_name:Contig22"
14103. 15350
/note="assembly_name:Contig23"
15451. 16663
/note="assembly_name:Contig24"
16764. 19156
/note="assembly_name:Contig25"
19257. 21199
/note="assembly_name:Contig26"
21300. 23500
/note="assembly_name:Contig27"
23601. 25776
/note="assembly_name:Contig28"
25877. 27788
/note="assembly_name:Contig29"
27889. 30449
/note="assembly_name:Contig30"
30550. 33166
/note="assembly_name:Contig31"
33267. 36506
/note="assembly_name:Contig32"
36607. 39573
/note="assembly_name:Contig33"
39674. 42099
/note="assembly_name:Contig34"
42200. 45279
/note="assembly_name:Contig35
vector_side:right"
45380. 48433
/note="assembly_name:Contig36"
48534. 51920
/note="assembly_name:Contig37"
52021. 55180
/note="assembly_name:Contig38"
55281. 57733
/note="assembly_name:Contig39"
57834. 60953

```

Query Match

1.5%; Score 20; DB 2; Length 185421;

Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 908 ctattttttttaagtaaa 927
|||||
Db 153950 CTATTGTTTAAAGTCAA 153931

RESULT 46
AC026160/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-1017A6 map 3p, WORKING DRAFT
AC026160
AC026160.2 GI:8121162
VERSION
KEYWORDS
SOURCE HTGS, HTGS-PHASE1, HTGS-DRAFT.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 198105)
Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
Dong, W., Fan, H., Fang, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,
Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G., Li, J., Li, L.,
Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y.,
Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y.,
Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R.,
Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,
Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H.,
Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y.,
Zhang, Z., Zhu, B., Yu, J. and Yang, H.
TITLE
Unpublished
Chromosome 3p genomic sequence
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 198105)
Kang, N., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,
Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,
Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,
Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D.,
Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J.
and Yang, H.
TITLE
Direct Submission
JOURNAL
Submitted (21-MAR-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P. R. China
On May 31, 2000 this sequence version replaced gi:7271977.
COMMENT
-----Genome Center
Center: Beijing Center
Center code: Beijing
Website: http://hgsc.igtp.ac.cn
http://www.genomics.org.cn
Contact: hgsc@igtp.ac.cn
-----Project Information
Center project name: 1% project
Center clone name: RP11-1017A6
-----Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 55% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 189600 bases at least Q40
Consensus quality: 194382 bases at least Q30
Consensus quality: 197736 bases at least Q20
Insert size: 190445; sum-of-contigs
Quality coverage: 4.76x in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
1
1195
1295
3027
3128
3128
4528
4628
8414
8514
13192
13292
13292
19448
19548
25716
25815
36292
36293
36393
49088
49187
49188
61855
61855
75325
75425
88685
88785
104411
104511
120476
120576
141215
141215
141315
1194: contig of 1194 bp in length
1294: gap of unknown length
3027: contig of 1733 bp in length
3127: gap of unknown length
3128: gap of unknown length
4527: contig of 1400 bp in length
4627: gap of unknown length
8413: contig of 3786 bp in length
8513: gap of unknown length
13191: contig of 4678 bp in length
13291: gap of unknown length
19447: contig of 6156 bp in length
19547: gap of unknown length
25715: contig of 6168 bp in length
25815: gap of unknown length
36292: contig of 10477 bp in length
36293: gap of unknown length
49087: contig of 12695 bp in length
49187: gap of unknown length
61854: contig of 12667 bp in length
61954: gap of unknown length
75324: contig of 13370 bp in length
75424: gap of unknown length
88685: contig of 13261 bp in length
88785: gap of unknown length
104410: contig of 15625 bp in length
104510: gap of unknown length
120475: contig of 15965 bp in length
120576: gap of unknown length
141214: contig of 20639 bp in length
141314: gap of unknown length
198105: contig of 56791 bp in length.
Location/Qualifiers
1. 198105
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-1017A6"
1. 1194
/note="assembly_name:Contig3"
1295. 3027
/note="assembly_name:Contig4"
3128. 4527
/note="assembly_name:Contig5"
4628. 8413
/note="assembly_name:Contig6"
8514. 13191
/note="assembly_name:Contig7"
13292. 19447
/note="assembly_name:Contig8
clone_end:SP6
vector_side:right"
19548. 25715
/note="assembly_name:Contig9"
25816. 36292
/note="assembly_name:Contig10"
36393. 45087
/note="assembly_name:Contig11"
49188. 61854
/note="assembly_name:Contig12
clone_end:T7
vector_side:right"
61955. 75324
/note="assembly_name:Contig13"
75425. 88685
/note="assembly_name:Contig14"
88786. 104410
/note="assembly_name:Contig15"
104511. 120475
/note="assembly_name:Contig16"
120576. 141214
/note="assembly_name:Contig17"
141315. 198105

```

BASE COUNT      62311 a 40661 c 38280 g 55347 t 1506 others
ORIGIN
/Note="assembly_name:Contig18"
Query Match      1.5%; Score 20; DB 2; Length 198105;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 525 catgaccatttgaact 544
Db 75312 CATGACAAATTGCTAACT 75293

RESULT 47
AC004673 236281 bp DNA PRI 18-NOV-1998
LOCUS      Homo sapiens Xp22-132-134 BAC GSHB-590J15 (Genome Systems Human BAC
DEFINITION 1library) complete sequence.
ACCESSION  AC004673
VERSION     AC004673.1 GI:3241934
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 236281)
AUTHORS     Muzny,D., Aronson,A.D., Adams,C., Brundage,E., Bunac,C.,
            Carvelli,K., Chacko,J., Chen,J., Di,W., Ding,Y., Dugan,S.,
            Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M.,
            Gorrell,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S.,
            Kampe,R., Karpathy,S., Kovar,C., Leal,B., Li,Y., Licharge,O.,
            Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L.,
            Rashid,N.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,
            Simon,M., Stovall,K., Tims,K.M., Todd,J., Vo,Q., Williamson,A.,
            Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
            Direct Submission
TITLE       Unpublished
JOURNAL     2 (bases 1 to 236281)
REFERENCE   Worley,K.C.
AUTHORS     Direct Submission
TITLE       Submitted (09-MAY-1998) Molecular and Human Genetics, Baylor
JOURNAL     College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 236281)
REFERENCE   Worley,K.C.
AUTHORS     Direct Submission
TITLE       Submitted (19-JUN-1998) Human Genome Sequencing Center, Department
JOURNAL     of Molecular and Human Genetics, Baylor College of Medicine, One
            4 (bases 1 to 236281)
REFERENCE   Worley,K.C.
AUTHORS     Direct Submission
TITLE       Submitted (18-NOV-1998) Human Genome Sequencing Center, Department
JOURNAL     of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 19, 1998 this sequence version replaced gi:221912.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.
            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.
            Sequence similarities were identified using PowerBlast by Jinghui
            Zhang.
            Exon/intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES
            Location/Qualifiers
            1..236281

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GSHB-590J15"
/chromosome="X"
/clone.lib="Genome Systems Human BAC library"
/map="Xp22-132-134"
complement(924..1306)
/rpt_family="MSTD"
complement(1502..1758)
/rpt_family="L1MA10"
1759..2712
/rpt_family="L1MB1"
2785..2990
/rpt_family="MIR"
complement(3221..3241)
/rpt_family="AT-rich"
complement(3349..3457)
/rpt_family="AT-rich"
complement(3598..4417)
/rpt_family="L1ME1"
4870..5127
/rpt_family="AluSp"
complement(5224..5262)
/rpt_family="AT-rich"
6244..6531
/rpt_family="AluSp"
6532..6704
/rpt_family="AluYb"
7136..8031
/rpt_family="L1M4"
8097..8133
/rpt_family="L2"
8132..8316
/rpt_family="L2"
8322..8773
/rpt_family="L2"
8928..9515
/rpt_family="L1M4"
9535..9832
/rpt_family="AluY"
complement(9942..10008)
/rpt_family="AT-rich"
complement(11483..11757)
/rpt_family="MER6"
complement(11761..12132)
/rpt_family="MER6"
11822..11970
/standard_name="AFM082w45, Chr. X, Homo sapiens"
complement(13900..13941)
/rpt_family="AT-rich"
14432..14729
/rpt_family="L1PA13"
14734..15008
/rpt_family="AluX"
15019..15538
/rpt_family="L1PA13"
15541..15840
/rpt_family="AluS4"
15841..18604
/rpt_family="L1PA13"
18605..18684
/rpt_family="(TAAA)n"
complement(19735..20039)
/rpt_family="AluS4"
complement(20188)
/rpt_family="L1M4"
complement(21141..21374)
/rpt_family="L1M4"
complement(21794..21850)
/rpt_family="L2"
22757..23176
/rpt_family="L1PA7"
23195..23238

```

Query Match	1.5%;	Score 20;	DB 9;	Length 236281
Best Local Similarity	100.0%;	Pred. No. 20;		

RESULT	48				
AC005529					
LOCUS	AC005529	318488 bp	DNA	PR1	01-JUN-2000
DEFINITION	Homo sapiens chromosome 22q12 clone bk256d12, complete sequence.				
ACCESSION	AC005529				
VERSION	AC005529.7	GI:4646249			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				

REFERENCE 3 (bases 1 to 318488)
AUTHORS Ying, F., Pan, H. and Roe, B.A.
TITLE Direct Substitution
Submitted (09-DEC-1998) Department of Chemistry and Biochemistry,
University of California, San Diego, La Jolla, CA 92037, USA
[Room 308 Newman

REFERENCE
AUTHORS
TITLE
JOURNAL

5 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (05-JAN-1999) Department of Chemistry and Biochemistry,
University of California, Room 708, Norman
Hall, University of California, 620 North

REFERENCE
7 (bases 1 to 318488)
AUTHORS
Ying, F., Pan, H. and Roe, B.A.
TITLE
Direct Submission
JOURNAL
Submitted (09-JAN-1999) Department of Chemistry and Biochemistry,
University of California, San Diego, La Jolla, CA 92037, U.S.A.
Accession Number 306

[illegible]

REFERENCE	10 (bases 1 to 318488)
AUTHORS	Ying, F., Pan, H. and Roe, B.A.
TITLE	Direct Submission

JOURNAL Submitted (16-JAN-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
11 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (19-JAN-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
12 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (24-JAN-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
13 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (18-FEB-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
14 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (27-FEB-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
15 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (04-MAR-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
16 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (22-APR-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
17 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (24-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
18 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (24-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
19 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (26-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
20 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (27-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
21 (bases 1 to 318488)
Ying, F., Pan, H. and Roe, B.A.
Direct Submission
Submitted (01-JUN-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

COMMENT
On Apr 22, 1999 this sequence version replaced g1:4417314. Because these overlapping clones came from different libraries there are numerous instances of insertions, deletions, and single nucleotide polymorphisms in the overlapping regions below.

AC000025(566c1) 38234 135405 (0) overlaps AC005529(BK256d12) 1 97163 (221325) AC000035(1147g11) 10813 38429 (0) overlaps AC005529(BK256d12) 1 27604 (290884) AC005529(BK256d12) 225 149508 (168980) overlaps AC005527(489c1) 1 149308 (0) AC005529(BK256d12) 72129 110244 (208244) overlaps AC002991(n47h6) 1 38140 (0) AC005529(BK256d12) 98641 180736 (137752) overlaps AC005894(BK294c2) 1 82091 (10).

FEATURES
Location/Qualifiers
1..318488
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22q12"
/clone="BK256d12"
BASE COUNT 82006 a 78473 c 79291 g 78718 t
ORIGIN

Query Match 1.5% Score 20: DB 9: Length 318488;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 711 tgaatggtcagagattga 730
|||||
Db 162190 TGAATGTCAGAGATTGA 162209

RESULT 49
G55641/c 496 bp DNA STS 30-MAR-2000
LOCUS SHGC-100838 Human Homo sapiens STS genomic, sequence tagged site.
DEFINITION G55641
ACCESSION G55641.1 GI:6120810
VERSION
KEYWORDS STS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
1 (bases 1 to 496)
Olivier, M. and Cox, D.R.
Unpublished, Olivier, M., Cox, D.R. (2000)
Unpublished (2000)

Contact: Michael Olivier, David R. Cox
Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801
Email: olivier@shgc.stanford.edu
Primer A: GTTGGATGGGAGACTTTCAGATG
Primer B: AGCGAACAATCATATACATCGTC
STS size: 282
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifig Gold Polymerase: 0.07 units/uL
Total Vol: 5 uL

Buffer: MG12: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

BAC ends sequenced at TIGR from the RPC11 BAC library. Designed

FEATURES and developed at the Stanford Human Genome Center.

SOURCE

Location/Qualifiers
1. .496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="1"
/clone_lib="Human"

STS

primer_bind

180. .461

primer_bind

180. .202

BASE COUNT

complement(439, .461)
181 a 72 c 112 g 131 t

ORIGIN

Query Match 1.4%; Score 19; DB 11; Length 496;

Best Local Similarity 100.0%; Pred. No. 43; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0;

QY 549 ctcacatatttcattgct 567

Db 232 CTCACATATTTCATTGCT 214

RESULT 50

AF217223 1087 bp DNA VRL 03-AUG-2000

LOCUS

Influenza B virus (B/Netherlands/580/89) hemagglutinin gene,

DEFINITION

partial cds.
AF217223
AF217223.1 GI:7650317

ACCESSION

AF217223.1 GI:7650317

KEYWORDS

Influenza B virus (B/Netherlands/580/89).

SOURCE

Influenza B virus (B/Netherlands/580/89).

REFERENCE

1 (bases 1 to 1087)
Osterhaus, A.D., Rimmelzwaan, G.F., Martina, B.E., Bestebroer, T.M. and

AUTHORS

Fouchier, R.A., Rimmelzwaan, G.F., Martina, B.E., Bestebroer, T.M. and

TITLE

Direct Submission
Submitted (20-DEC-1999) Virology, Erasmus University, Dr.

JOURNAL

Molewaterplein 50, Rotterdam 3015 GE, Netherlands

FEATURES

Location/Qualifiers
1. .1087
/organism="Influenza B virus (B/Netherlands/580/89)"
/strain="B/Netherlands/580/89"

CDS

<1..>1087
/note="HA1 domain"

/codon_start=2
/product="hemagglutinin"

/protein_id="AAF65990.1"

/db_xref="GI:7650318"

/translation="MKATIVLMVVTNSNADRICGTGTTSSSPHVYKTAOGENVYTV
IPITPTPKSHRANLKGKTRGKLCPCNCLDIDVAIRPKVCVGTIPSAKASIIHAY
RPVTSQCFPIIMHDRIKIRDPNLRLGYENIRLSTQVINAERKPGPIRLGTSGCPN
VTSRNGEFATMAVPRDKTAIINPLTVEVPYICNGEDQITVWGFHSNDKTKOMKLY
GDSNPKFTSSANGVTTHVVSQIGFPPNTEGGLPQSGRIIVDVWVOKPGKGTIVY
ORGVLLPKQWMCAGRSKIVKSLPLIGADCLHEKXGGLNKSFPYTGEGAIAIGNC
PIWKTPIKLANGTKYRPPAKLIKRGF"

BASE COUNT

381 a 243 c 227 g 236 t

ORIGIN

Query Match 1.4%; Score 19; DB 14; Length 1087;

Best Local Similarity 100.0%; Pred. No. 47; Mismatches 0; Indels 0; Gaps 0;

Matches

19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 764 tgaaccccccaacagtaa 782

Db 621 TGAACCCCCCAACAGTAA 603

Search completed: May 2, 2002, 13:18:43
Job time: 9727 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2002, 07:45:15 ; Search time 28.73 Seconds
(without alignments)
2306.348 Million cell updates/sec

Title: US-09-645-192-2

Perfect score: 2389
Sequence: 1 MKIFKCYFHTLQOKVFILF.....DWITLPSKELFMDRLTTTS 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_17:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhcc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.vertebrate:*
15: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2389	100.0	453	4 Q9P109	Q9P109 homo sapien
2	868	36.3	438	4 O95395	O95395 homo sapien
3	866.5	36.3	428	11 O64165	O64165 rattus norv
4	848	35.5	440	12 Q912K2	Q912K2 bovine herp
5	847	35.5	398	11 Q9D8A3	Q9D8A3 mus musculu
6	844.5	35.3	428	11 O35981	O35981 mus musculu
7	842	35.2	440	12 Q99CM3	Q99CM3 bovine herp
8	645.5	27.0	356	11 Q9D2A8	Q9D2A8 mus musculu
9	508.5	21.3	314	4 Q9HCY8	Q9HCY8 homo sapien
10	326	13.6	445	5 Q19730	Q19730 caenorhabdi
11	323	13.5	425	5 O02234	O02234 caenorhabdi
12	315.5	13.2	454	5 Q19729	Q19729 caenorhabdi
13	296	12.4	401	5 O02314	O02314 caenorhabdi
14	294	12.3	472	5 Q20406	Q20406 caenorhabdi
15	275	11.5	402	5 O02315	O02315 caenorhabdi
16	245	10.3	459	5 O02309	O02309 caenorhabdi
17	231	9.7	876	5 Q9W034	Q9W034 dirosophila
18	229.5	9.6	454	5 O45866	O45866 caenorhabdi
19	224.5	9.4	864	11 Q9EP10	Q9EP10 rattus norv

20	222.5	9.3	322	2 Q9A4H4	Q9A4H4 caulobacter
21	220.5	9.2	489	5 Q21796	Q21796 caenorhabdi
22	219.5	9.2	470	5 O16776	O16776 caenorhabdi
23	219.5	9.2	486	5 Q93643	Q93643 caenorhabdi
24	219.5	9.2	865	4 Q9H1B5	Q9H1B5 homo sapien
25	214.5	9.0	616	11 Q9EP10	Q9EP10 mus musculu
26	213.5	8.9	821	11 Q9EP11	Q9EP11 rattus norv
27	213.5	8.9	787	4 Q9H1B6	Q9H1B6 homo sapien
28	211.5	8.9	829	11 Q9EP11	Q9EP11 mus musculu
29	203.5	8.5	472	5 Q26334	Q26334 caenorhabdi
30	202.5	8.5	753	5 O02312	O02312 caenorhabdi
31	197.5	8.3	448	10 Q9Z0Z7	Q9Z0Z7 arabidopsis
32	194.5	8.1	513	5 Q9VFP0	Q9VFP0 caenorhabdi
33	193.5	8.1	478	5 Q9XXL2	Q9XXL2 caenorhabdi
34	188	7.9	467	5 Q9U3H9	Q9U3H9 caenorhabdi
35	187	7.8	447	10 Q9LR71	Q9LR71 arabidopsis
36	182.5	7.6	513	5 Q9XTM4	Q9XTM4 caenorhabdi
37	181	7.5	424	10 Q9LFE0	Q9LFE0 arabidopsis
38	178	7.5	367	5 Q22481	Q22481 caenorhabdi
39	175.5	7.3	406	10 Q9LNN5	Q9LNN5 arabidopsis
40	172.5	7.2	378	10 Q9S6S9	Q9S6S9 arabidopsis
41	169.5	7.1	378	10 Q9ASZ7	Q9ASZ7 arabidopsis
42	168	7.0	395	10 Q9C9A1	Q9C9A1 arabidopsis
43	166.5	7.0	434	10 Q9LFD0	Q9LFD0 arabidopsis
44	151	6.3	447	10 Q9FLD7	Q9FLD7 arabidopsis
45	139	5.8	402	10 Q9AR03	Q9AR03 oryza sativ

ALIGNMENTS

RESULT 1
ID Q9P109 PRELIMINARY; PRT: 453 AA.
AC Q9P109;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CORE 2 BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE 3.
GN C2GN73.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20219156; PubMed-10753916;
RA Schwiennek T., Yen J.C., Levery S.B., Keck B., Merix G.,
RT van Kessel A.G., Fukuda M., Clausen H.;
RT "Control of O-glycan branch formation. Molecular cloning and
RT characterization of a novel thymus-associated core 2 beta1,6-N-
RT acetylglucosaminyltransferase.";
RT J. Biol. Chem. 275:11106-11113(2000).
DR EMBL: AF132035; AAF63156.1;
DR InterPro: IPR003406; Branch.
DR Pfam: PF02485; Branch.1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 453 AA; 53052 MW; B43794D4427F41CA CRC64;

Query Match 100.0%; Score 2389; DB 4; Length 453;
Best Local Similarity 100.0%; Pred. No. 1e-173;
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKIFKCYFHTLQOKVFILFMTLSLKLNVRLFPQKDIYVERSLSTSPVRRNY 60
DB 1 MKIFKCYFHTLQOKVFILFMTLSLKLNVRLFPQKDIYVERSLSTSPVRRNY 60
QY 61 THVKDVRREVNCSGIVDEPLEIGKSLERRDIILEDVDVWMTSCDIYQTLRGTA 120
DB 61 THVKDVRREVNCSGIVDEPLEIGKSLERRDIILEDVDVWMTSCDIYQTLRGTA 120
QY 121 QKLVSEKSFPIAYSLVYHKDAIWERLIIAIYNOHNIYCIHYDRKAPDFKVAAMNLA 180

```

DB 121 OQLSKSEKSPFLAYSLVVKDAIMVERLHAITNOHTICIHDKRAPDTFYAMNNLA 180
QY 181 KCFNSIFIAKLEAVEVAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELV 240
DB 181 KCFNSIFIAKLEAVEVAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELV 240
QY 241 SELKINGANMLETVKPPNSKLEFTHHLLRPVYEVKLPRTINISKAPPHNIOIFV 300
DB 241 SELKINGANMLETVKPPNSKLEFTHHLLRPVYEVKLPRTINISKAPPHNIOIFV 300
QY 301 GSAFVLISQAFVVKYIFENNSIQDFEAFNSKDTYSPDEHFWATLLIRVGPISPEISRSADVS 360
DB 301 GSAFVLISQAFVVKYIFENNSIQDFEAFNSKDTYSPDEHFWATLLIRVGPISPEISRSADVS 360
QY 361 DLQSTRLVKNWYIEGFYPSCTGSHLSVCIYGAELRMLIKDGHFANKFDSKVPDL 420
DB 361 DLQSTRLVKNWYIEGFYPSCTGSHLSVCIYGAELRMLIKDGHFANKFDSKVPDL 420
QY 421 IKCLAEKLEEQORDWITLPSSEKLFMDRLITTS 453
DB 421 IKCLAEKLEEQORDWITLPSSEKLFMDRLITTS 453

```

RESULT 2
ID 095395 PRELIMINARY; PRT; 438 AA.

```

AC 095395;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE.
GN C2/4GN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RX MEDLINE=9915671; Pubmed=9915662;
RC TISSUE=BRIN;
RA Yeh J.C., Ong E., Fukuda M.;
RT "Molecular cloning and expression of a novel beta-1, 6-N-
RT acetylglucosaminyltransferase that forms core 2, core 4, and I
RT branches.";
RL J. Biol. Chem. 274:3215-3221(1999).
RP SEQUENCE FROM N.A.
RX MEDLINE=99143102; Pubmed=9986882;
RA Schiavetek T., Nomoto M., Levery S.B., Merx G., van Kessel A.G.,
RT "Control of O-glycan branch formation. Molecular cloning of human CDNA
RT encoding a novel beta1,6-N-acetylglucosaminyltransferase forming core
RT 2 and core 4.";
RL J. Biol. Chem. 274:4504-4512(1999).
DR EMBL: AF102542; AAD10824.1;
DR EMBL: AF038650; AAD21525.1;
DR InterPro: IPR003406; Branch.
DR Pfam: PF02485; Branch: 1.
DR Transferrase: Glycosyltransferase.
SQ SEQUENCE 438 AA; 50863 KM; 1FF0A/B451C88407 CRC64;

```

Query Match 36.3%; Score 868; DB 4; Length 438;
Best Local Similarity 43.0%; Pred. No. 4,6e-58;
Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;

```

QY 23 LMLSLKLKLN-----KRLPQKITYVESLSISFVRN-RTHVADERYVNSCG 75
DB 13 LMLLGYMLATVALKLSFRLKCDSDHLESRSSQYCRNLINFLKLPKRSINCSG 72
QY 76 IV--EDEL--EIGKSLERIRDIIDLEDDVVAAMSDCDIYQTLRGYAKIVSEKSF 131
DB 76 IV--EDEL--EIGKSLERIRDIIDLEDDVVAAMSDCDIYQTLRGYAKIVSEKSF 131

```

```

DB 73 VTRGDEAVLQALINNLVYKKR--EPFTDHYLSLTDCEHKAERKFIQEPLESKEVEF 131
QY 132 PLAYSIVYKDAIMVERLHAITNOHTICIHDKRAPDTFYAMNNLAKCFNSIFIAK 191
DB 132 PLAYSIVYKDAIMVERLHAITNOHTICIHDKRAPDTFYAMNNLAKCFNSIFIAK 191
QY 192 LEAVEVAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELKINGANM 251
DB 192 LEAVEVAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELKINGANM 251
QY 252 LTVKPPNSKLEFTHHLLRPVYEVKLPRTINISKAPPHNIOIFGSAFVLISQAF 311
DB 252 LTVKPPNSKLEFTHHLLRPVYEVKLPRTINISKAPPHNIOIFGSAFVLISQAF 311
QY 312 VKYIFENNSIQDFEAFNSKDTYSPDEHFWATLLIRVGPISPEISRSADVS 370
DB 312 VKYIFENNSIQDFEAFNSKDTYSPDEHFWATLLIRVGPISPEISRSADVS 370
QY 366 WOCHGEDIDKGAAPACSGIHORALICVYAGDLNMLQNHILLANKFDPKVDNALQCLE 425
DB 366 WOCHGEDIDKGAAPACSGIHORALICVYAGDLNMLQNHILLANKFDPKVDNALQCLE 425
QY 426 EKL 428
DB 426 EKL 428

```

RESULT 3
ID 064165 PRELIMINARY; PRT; 428 AA.

```

AC 064165;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ENZYMAIC GLYCOSYLATION-REGULATING GENE PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RX MEDLINE=96013636; Pubmed=7560067;
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HEART;
RA Nishio Y., Warren C.E., Buczek-Thomas J.A., Rulis J., Koya D.,
RT "Identification and characterization of a gene regulating enzymatic
RT glycosylation which is induced by diabetes and hyperglycemia
RT specifically in rat cardiac tissue.";
RL J. Clin. Invest. 96:1159-1167(1995).
DR EMBL: S79797; AAB35697.2;
DR InterPro: IPR003406; Branch.
DR Pfam: PF02485; Branch: 1.
SQ SEQUENCE 428 AA; 49826 MW; AE25A18172897AA8 CRC64;

```

Query Match 36.3%; Score 866.5; DB 11; Length 428;
Best Local Similarity 42.3%; Pred. No. 5,8e-58;
Matches 183; Conservative 73; Mismatches 136; Indels 41; Gaps 11;

```

QY 28 LKLINVRLEF--PQNDIVLV-EYSLSTSPFVRN-----YTHVK---DEVRYEYVNS 74
DB 1 MLRNLFRRLFLSYPTKTYFNVVLSTLTFGVRHOKPEVSVSHLESLSDDDNSNNVCT 60
QY 75 GIVQDEPLEIK-SLEI-----RRDIIIDDDVVAAMSDCDIYQTLRGYAKIV 124
DB 61 KYLGDEPEIQKLEILITVQFKRRPR-----TPHDYINMRDCASFIRIRKTYMEPL 114
QY 125 SKEKSPFIAYSLVVKDAIMVERLHAITNOHTICIHDKRAPDTFYAMNNLAKCFNS 184
DB 115 TKEKSPFIAYSLVVKDAIMVERLHAITNOHTICIHDKRAPDTFYAMNNLAKCFNS 174
QY 185 NIFIAKLEAVEVAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELK 244
DB 185 NIFIAKLEAVEVAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELK 244

```


Db 175 NFEVASQLESVYASWSRYKADINCKMDLYRMNANKYLINLGMDFPIKTNLEIYRKAK 234
 QY 245 KUNGAMLETVRPNSKLEBFTYHHLRVPYEVYKLTPTRTNISKAPENHICIFVGSAY 304
 Db 235 SFTGENSLETERKPPNKEENK-----KRTYVDGKL-INTGVYKQPLKPLTFBGSAY 288
 QY 305 FVLSQAFVKYIFNNSIVODFANSKDTYSPDEHFMTLRVPGICPEISRSAO-DVSDIQ 363
 Db 289 FVYTRREYGVLENKKNIKQKFMWAQDTYSPDELFMTIORIPPEVPGSLPSHRYDLSDMN 348
 QY 364 SKRTIKWNYEGEF-----YPSCTGSHLRVSCVIGAAELRMLIKDGHWFANKFDSKVPD 418
 Db 349 AVARFKWQYFEGDVSNGADYPPGSGVHVASCVGVGDLSSMLKRHHFFANKFDMVDVP 408
 QY 419 ILIKLAEKLEEQ 431
 Db 409 FALQCEHLRHK 421

RESULT 4
 Q91ZK2 PRELIMINARY; PRT; 440 AA.
 ID Q91ZK2
 AC Q91ZK2
 DT 01-OCT-2000 (TEMBLrel. 15, Created)
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE.
 GN BOREF3-4.
 OS Bovine herpesvirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 OC NCBI_TaxID=10385;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=V. TEST;
 RX MEDLINE=20283898; PubMed=10811884;
 RA Vanderplasschen A., Markine-Gorlaynoff N., Lomonte P., Suzuki M.,
 RA Hiraoka N., Yeh J.-C., Bureau F., Williams L., Thiry E., Fukuda M.,
 RA Pastoret P.-P.;
 RT "A multipotential beta-1,6-N-acetylglucosaminyl-transferase is encoded
 RT by Bovine herpesvirus type 4."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5756-5761(2000).
 DR EMBL; AF231105; AAF72001.1;
 DR InterPro; IPR003406; Branch.
 DR Pfam; PF02485; Branch; 1.
 KW Transferase; Glycosyltransferase.
 SQ SEQUENCE 440 AA; 50701 MW; 71B9C3B6B4A949BD CRC64;

Query Match 35.5%; Score 848; DB 12; Length 440;
 Best Local Similarity 44.0%; Pred. No. 1.5e-56;
 Matches 168; Conservative 72; Mismatches 124; Indels 18; Gaps 6;
 QY 60 YTHVDEVREYVNCSGIY--EOEPL--EIGKSLERRDIDLEDDDVVAMTSCDIYOT 115
 Db 60 YNSLKLPAKRSINCIGIRGDEAVYQALDNLVEYKKR-PLPDTYLLNTIROCERKA 118
 QY 116 LRGVQKLVSEKESFPIAYSLVHKDAIMVERLTHAIYNQHNITCIHYDKRAPDFEYVA 175
 Db 119 QKRFIOPLSKBELDFPIAYSMVYHEKLENERLRAVYAQONICVHADVKSPEPFKEA 178
 QY 176 MNNLAKCFSNIFASKLEAVEYAHISRLQADINCLSDLSKSIOKRYVINLGGDFPLKS 235
 Db 179 VVAITSCEFPNVMAASKLPVYVYASMSRVQADLNCMEDLDSQSVSKYLLNTICGIDFPK 238
 QY 236 NFEVLSKIKLNGANMLETVPKNSKLEBFTYHHLRVPYEVYKLTPTRTNISKAPPHN 295
 Db 239 NAEVYVATLKLKLGKXMSMESEVPSKKNMKRYEVTDLTPTSTKI-----KDPDPDN 291
 QY 296 IOIFVGSAYEVLSQAFVKYIFNNSIVODFANSKDTYSPDEHFMTLRVPGICPEI-SR 354
 Db 292 LPEFTGNATFVASRAFYAVHLDNPKRSQILVENVKOTYSPDEHLMTLQGRAPMPSVSH 351

QY 355 SAQDVSDLSQSTRVLYKWNYYEGF-----YPSCTGSHLRVSCVIGAAELRMLIKDGHWF 409
 Db 352 PKYHSDMTALATRLYKWOYHEGDVSMGAPVAPCSGIRHRAICIGADLWILQNHLLA 411
 QY 410 NKFDKVPILIKLAEKLEEQ 431
 Db 412 NKFDPRVDNVLQCLEEYLRHK 433

RESULT 5
 Q9D8A3 PRELIMINARY; PRT; 398 AA.
 ID Q9D8A3
 AC Q9D8A3
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE 2010013H22RIK PROTEIN.
 GN 2010013H22RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arawaka T., Hara A., Fukunishi Y., Kono H., Adachi J., Yamanaka I.,
 RA Aizawa K., Izawa M., Niehi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.O., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
 RA Wuzhuhaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
 RA Hayashiraki Y.,
 RT "Functional annotation of a full-length mouse CDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK008234; BAB25548.1;
 DR MGD; MGI:1919327; 2010013H22RIK.
 DR InterPro; IPR003406; Branch.
 DR Pfam; PF02485; Branch; 1.
 SQ SEQUENCE 398 AA; 46000 MW; 275ABE93909CC4B6 CRC64;

Query Match 35.5%; Score 847; DB 11; Length 398;
 Best Local Similarity 42.4%; Pred. No. 1.6e-56;
 Matches 162; Conservative 79; Mismatches 123; Indels 18; Gaps 5;

QY 60 YTHKXDEVREYVNCSGIYEOEPLCK-----SLFIRRDIDLEDDDVVAMTSCDIYOT 115
 Db 18 YTKTLPAKRSINCISGIRGEQKAVYQALDNLNLEIKKQQL-FTEDLYLMTADCHFT 76
 QY 116 LRGVQKLVSEKESFPIAYSLVHKDAIMVERLTHAIYNQHNITCIHYDKRAPDFEYVA 175
 Db 179 VVAITSCEFPNVMAASKLPVYVYASMSRVQADLNCMEDLDSQSVSKYLLNTICGIDFPK 238
 QY 236 NFEVLSKIKLNGANMLETVPKNSKLEBFTYHHLRVPYEVYKLTPTRTNISKAPPHN 295
 Db 239 NAEVYVATLKLKLGKXMSMESEVPSKKNMKRYEVTDLTPTSTKI-----MTSKRRTPPNN 291
 QY 296 IOIFVGSAYEVLSQAFVKYIFNNSIVODFANSKDTYSPDEHFMTLRVPGICPEI-SR 354
 Db 292 LPEFTGNATFVASRAFYAVHLDNPKRSQILVENVKOTYSPDEHLMTLQGRAPMPSVSH 351

```

QY 296 IQIFVGSAYFVLSQAFKYYIFNNSIVODFFAMSKDTSPDEHFWATLIRVPGIGELISRS 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 LTMFGNAYVASRDEIFEHVSNSKARQILEMVKDYSPEDEHFWATLIRASMGSPDLH 309
QY 356 AQ-DVSDQSKTRLYKKNYYEGFF-----YPSCTGSHLRSCVCIIGAELRMLIDGMFPA 409
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 RKFDMSDKRAIRALRLTKWDEHEDENGAPYTSCSGIIHORAVCVSGDLHWILONHLLA 369
QY 410 NKFDKVDPIILIKLAEKLEQ 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 NKFDKVDNVLQCLEEYLRHK 391

RESULT 6
ID 035981 PRELIMINARY; PRT; 428 AA.
AC 035981;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE GLUCOSAMINYL (N-ACETYL) TRANSFERASE 1, CORE 2 (BETA-1,6-N-
DE ACETYLGLUCOSAMINYLTRANSFERASE).
GN GCNT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C; TISSUE=SUBMAXILLARY GLAND, AND KIDNEY;
RX MEDLINE=98001705; Pubmed=9341170;
RA Sekine M., Nara K., Suzuki A.;
RT "Tissue-specific regulation of mouse core 2 beta-1,6-N-
RT acetylglucosaminyltransferase."
RL J Biol Chem 272:27246-27252(1997).
DR EMBL; D87333; BAA22899.1; -
DR MGI; D87332; BAA22898.1; -
DR InterPro; IPR003406; Branch.
DR Pfam; PF02485; Branch; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 428 AA; 49831 MW; F0D10A6ADB23B92C CRC64;

Query Match 35.3%; Score 844.5; DB 11; Length 428;
Best Local Similarity 42.4%; Pred. No. 2.7e-56;
Matches 183; Conservative 71; Mismatches 139; Indels 39; Gaps 11;

QY 28 LKILNVRRLF--POKDIY-LVEYSLSTSPFYRNRYTHVKDE---VRY-----EV 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MERNLFRRLFCSPKTYFMLVLSLTFSVLR---IHQKPEFVSRRHLELAGDDPYSNV 57
QY 72 NCSGIYEDPLEIGK-----SLEIRRRDIDLEDDVVAWTSDDCIQYTLRGVAKLVLS 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 NCKTILQGGPEEFQKVKLELTVQFKRP--RMTPHDYINMTRDCAASFIRTRKYIVEYEL 115
QY 126 KEKSPFIYSLVHKDAMVERLLHAITNCHNITCIHDKRAPTFFVAMNKLAKCSN 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 KEKVGPIAYSIYVHKRIEMDLRLAIYMPQNFYCIHVDRAESFSFLAAVGASCDN 175
QY 186 IFIASLAEVFAHISRLQADLNCLSLQKYYINLCGQPFLEKSNELSELEK 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 VFVASQLESVYASMSRVAADLNCMKDLYRMANKKYLNLGQMPFIKTNLEIYRKLK 235
QY 246 LNCANMLETVKPPNSKLEFTYHHELRVYEVYKLPITNISKAPRNIIQIFGSAVF 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 STANNLETEKMPNKEERWK-----KRAYVVDGKL-TNTGIYKAPPLKPTLFGSAYF 289
QY 306 VLSQAFVKTIFNNSIVODFFAMSKDTSPDEHFWATLIRVPGIGELISRS 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 VYTRVYGVLENENIKLMAKADPTYSPEDEHFWATLIRVPGIGELISRS 349
QY 365 KTRLYKKNYYEGFF-----YPSCTGSHLRSCVCIIGAELRMLIKDGMFPAKFSKVDPI 419

```

```

Db 350 IARFVWQYFEGDVNSGAYPPOSGYHVSVCVFGAGDLSWMLRQHHLEFANKEDIDVDPF 409
QY 420 LIRKLAKELEQ 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 AIOCDDEHLRRK 421

RESULT 7
ID 099CW3 PRELIMINARY; PRT; 440 AA.
AC 099CW3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VIRAL BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE.
OS Bovine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=10385;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20583805; Pubmed=11152491;
RA Zimmermann W., Broli H., Ehlers B., Buhk H.-J., Rosenthal A.,
RA Goltz M.;
RT "Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and
RT Identification of an Origin of DNA Replication."
RL J Virol 75:1186-1194(2001).
DR EMBL; AF318573; AA07999.1; -
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 440 AA; 50763 MW; C8BB3EAB8FDBDD9 CRC64;

Query Match 35.2%; Score 842; DB 12; Length 440;
Best Local Similarity 43.7%; Pred. No. 4.4e-56;
Matches 167; Conservative 72; Mismatches 125; Indels 18; Gaps 6;

QY 60 YTHVKEVRYEVNCSGIY--EQEPL--EIGKSLERRRDIDLEDDVVAWTSDDCIYOT 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 YNSIKLPAKKSINCSGITRQDQAVQALDNEVKKR--SPLTGYVINIRDCERFRA 118
QY 116 LRGAQKLVKKEKSPFIAYSLVHKDAMVERLLHAITNCHNITCIHDKRAPTFFKVA 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 QKRTQPLSKRELDPIFAISWVHEKTEFDELLRAVAPQNTCYHVDVSPETFKRA 178
QY 176 MNNAKCFNSIFLASKLEAVEYAHISRLQADLNCLSLQKYYINLCGQPFLEK 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 VKATISCFPNVMAKSLVPVYASWSRVQADLNCMEDLQSSVPMKYLINTCGTDEPIKT 238
QY 236 NFELVSELKLNQANMLETVKPPNSKLEFTYHHELRVYEVYKLPITNISKAPRNH 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 NAEVNLAKMLKGNMSSEVSPSESKNNRKYRIEYVDLTPYSKM-----KDPDPDN 291
QY 296 IQIFVGSAYFVLSQAFKYYIFNNSIVODFFAMSKDTSPDEHFWATLIRVPGIGELISRS 354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 LPMFTGNAYFASRAFYQHLNDNPKSQLVEMWKDYSPEDEHFWATLIRASMGSPVSH 351
QY 355 SAQVSDQSKTRLYKKNYYEGFF-----YPSCTGSHLRSCVCIIGAELRMLIKDGMFPA 409
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 PKYHISDMTALARLVKQYHGGVSMGAPAPCSGIIHRAICITYGADLTIWILONHLLA 411
QY 410 NKFDKVDPIILIKLAEKLEQ 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 NKFDKVDNVLQCLEEYLRHK 433

RESULT 8
ID 09D2A8 PRELIMINARY; PRT; 356 AA.
AC 09D2A8;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

```

DE 5330430K10RK PROTEIN.
 GN 5330430K10RK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=PIUTITARY GLAND;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Akakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kodota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Riechmann W., Gaasterland T., Gissi C., King B., Kochia H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirni L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Monbets P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL NCBI:409:685-690(2001).
 DR EMBL: AK019924; BAB31918.1; -
 DR MGD: MGI:1925531; 5330430K10RK.
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch: 1.
 SQ SEQUENCE 356 AA; 40446 MW; 0D67F48615656FD6 CRC64;

Query Match 27.0%; Score 645.5; DB 11; Length 356;
 Best Local Similarity 45.2%; Pred. NO. 2.9e-41; Indels 7; Gaps 3;
 Matches 132; Conservative 40; Mismatches 113;

QY 110 CDIVOTLNGVAKIVSKREKSPFIAYSLVHKDAIMVERLIHAIYNOHNYCIHYDRKAP 169
 DB 72 CPQYRQSHYITSPLESEEAFAFLAYIVIKHDETFERLPRAIYPOQNYCVAVHDSKAT 131
 QY 170 DTFKXANNLAKCFNSNITLKLAEVFAHISRLQADLNCSDLSKSIQWKYVINLCCG 229
 DB 132 DTFKXAVROLSCFPMNAELASKEGVYVGGFSLQADLNCMDLVAASKVPKYLNLTCGQ 191
 QY 230 DFLKSNFELVSELKLGANLLETKPPNSKLERFTYHHLRVRVEYVLPRTNTISK 289
 DB 192 DEPLKTNKEIINHKLKRGKNTPEGLPRAIYVATKYVHDERKDKDY--FMKNTNLK 249
 QY 290 EAPPNIOIFNGSAFVLSQAFVKYTFNNISYODFFANSKDTYSPDEHFWATLIRVGP 349
 DB 250 TPPPQOLITTFCTAYVALTRDFVNLIDERAIALLEMSKDTYSPDEHFWATLIRVGP 309
 QY 350 GEISSADVDSDLSKTRFLVKNMYEGFVPSCTGSHLRSCVICAELRWL 401
 DB 310 GSMFPAASWTGNL-----RAYKMDMEA-KHGCGCHKVRSAGATLPVADPRDL 356

RESULT 9
 Q9HCY8 PRELIMINARY; PRT; 314 AA.
 Q9HCY8:
 DT 01-MAR-2001 (TREMREL. 16, Created)
 DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE D1153D9.2 (A NOVEL PROTEIN SIMILAR TO BETA 1,6-N-
 DE ACETYLGALACTOSAMINYLTRANSFERASE.) (FRAGMENT).
 GN D1153D9.2.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Clark G.;
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL109806; CAC03741.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch: 1.
 FT NON-TER 1
 FT NON-TER 1
 FT NON-TER 1
 SQ SEQUENCE 314 AA; 36042 MW; 8E1B3E3976B2EE85 CRC64;

Query Match 21.3%; Score 508.5; DB 4; Length 314;
 Best Local Similarity 44.0%; Pred. NO. 6.4e-31;
 Matches 99; Conservative 41; Mismatches 82; Indels 3; Gaps 3;

QY 124 VSKEKSPFIAYSLVHKDAIMVERLIHAIYNOHNYCIHYDRKAPDFKXANNLAKCF 183
 DB 2 LSAEEDSFSLAYITTHKEKELAMEVQLRAIYPOQNYCIHVDKAPMKYKTAVQTLVNCF 61
 QY 184 SNFISKLEAVYFAHISRLQADLNCSDLSKSIQWKYVINLCCGDFLKNFELVSL 243
 DB 62 ENFISKTEKVAAYAGFTRLQADLNCMDLVAASKVPKYLNLTCGQDFPKTNRELIHI 121
 QY 244 K-KLNGANLLETKPPNSKLERFTYHHLRVRVEYVLPRTNTISKAPPNIOIFNGS 302
 DB 122 RSKMSDKNIRPGVIOPLHLSKTSQSHLEFVFKSGIYAP--NNRFXKDPKPNLITVFGS 179
 QY 303 AYVLSQAFYKVFNNISYODFFANSKDTYSPDEHFWATLIRVGP 347
 DB 180 AYVLSQAFYKVFNNISYODFFANSKDTYSPDEHFWATLIRVGP 224

RESULT 10
 Q19730 PRELIMINARY; PRT; 445 AA.
 AC Q19730;
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE F32D6.11 PROTEIN.
 GN F32D6.11.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilkinson J.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurry A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkhen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans".
 RL Nature 368:32-38(1994).
 DR EMBL: Z71262; CAA95817.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch: 1.
 SQ SEQUENCE 445 AA; 51330 MW; 4380A86BAF05DC6D CRC64;

Matches 103: Conservative 83; Mismatches 184; Indels 102; Gaps 17;

QY 32 LNWRR-----LEPQDIYL-----VEYSLSTSPFVNRHYHVDVRYVNGSGIYE 78
 Db 24 MNVRRFKISOFIYSSFTYIVLRIGFVDYL-----VPENRK-----LE 63
 QY 79 QEPLEIKSLSE--IRRDI-----IDLEDDVYAMTSDCDIYOTLRGVAOKIVSKRE 127
 Db 64 NIPLECSVINGTENRKRKISARQMDHFDVVEHEINSSKNVCSTIDKYINFRIPSSRE 123
 QY 128 EKSPFIAYSLVNHDAIMVERLIIHAIYNQNIYCIHYDRAPDTEKYAMNNLAKCFSN-- 185
 Db 124 EAEPFLANGIYVYKTIYQVLTQMSLFYQPOHMFICITVDQSPNEKYKSVIALPSCEPNMH 183
 QY 186 IFISKLEAVEXAHISRLQADNLCLSLKSIQWKYVINLGGODEPLKSNFELYSCLK 245
 Db 184 VFIG---EPSCWGSFGIKKNYITCFNWLKSKQKWKYQVLSIGDLEIRNLKLVRFKA 240
 QY 246 LGANMLETVPKPSKLERFTYHHLRVRVYEVYKLPRTNISKADPPHNIQIFVGSAYF 305
 Db 241 LNGS-----NMVDVSTF---EVDRY-----KMEGVLP--MPYKSSMSY 276
 QY 306 VLSCAFVYIINNSIYQDFAMSKDTYSPDEHFMTLIRVPG---IPGEISRSADQVSD 362
 Db 277 VVPRGADYLLISPRVQKLKYLKTLWLPDESFWSTYVGSFALLPVGST--RYRDLIML 334
 QY 363 QSKTRL-----VKMNYEFG---FYPSCTGSHLSRVCITYGAELFWLIKDGHW 408
 Db 335 RKNFLRPFREYTVNSITSYIGRYQVGMQKECFGKVKYDSCYGVEDIEMITRELY 394
 QY 409 ANKFESKVDPLIKCLAEKLEEQ-----RDWITLSEKLFMDRNT 450
 Db 395 AHKYLEFQPAFMCKLEKVRRLRSISPDALFSAFSISQMPTELYQCKAT 446

RESULT 15
 002315 PRELIMINARY; PRT; 402 AA.

AC 002315.
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE T15D6.3 PROTEIN.
 GN T15D6.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dobson R.;
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans";
 RL Nature 368:32-38(1994).
 DR EMBL: Z83125; CAB05621.1;
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 SQ SEQUENCE 402 AA; 46135 MW; F6AD721976703307 CRC64;

Query Match 11.5%; Score 275; DB 5; Length 402;
 Best Local Similarity 24.0%; Pred. No. 5, 1e-13;
 Matches 101: Conservative 69; Mismatches 169; Indels 82; Gaps 14;

QY 16 VFLEFLTWLSSLKLLNVRLLFPQKDIYLVYSTSPFVNRHYHVDVRYVNGSGIYE 75
 Db 19 IFLLFAKALIPRLLENPNQRIHPPTDL-----DDL--QINCTN 56
 QY 76 IQEERLEIKSLSEIRRDIDLEDDVYAMTSDCDIYOTLRGVAOKIVSKREKSPFIAY 135
 Db 57 ILOGFNE--NLELINTMT---ENKMNSTDRQITLSMRFKAYLSEEAFAFPLSE 111
 QY 136 SLVNRDAIMVERLIIHAIYNQNIYCIHYDRAPDTEKYAMNNLAKCFSNFIASKLAV 195
 Db 112 GLIVYKELSQVLFLLSSIIYQPOHMFICITVDQSPNEKYKSVIALPSCEPNMH 166
 QY 196 EYAHISRLQADNLCLSLKSIQWKYVINLGGODEPLKSNFELYSCLKLNGANMLET 255
 Db 167 KWSGFEIINSYGCLEFLSHLSKSDMKYFQYLSGVDIPKTNLEMYRIKLRLNG 221
 QY 256 KPNKSLERFTYHHLRVRVYEVYKLPRTNISKADPPHNIQIFVGSAYF 315
 Db 222 NIGIS-----TYEDRL-----LNGKNKTESP--LPLFSSLSLIPRAANT 262
 QY 316 FNNSIYQDFAMSKDTYSPDEHFMT-----LIRVPGIPGEISRSADQVSDQSKTRLY 369
 Db 263 SSSSVPOQLLEFLRTTVADEGFWGLFGNKDLEFNVPGSFNNPNDPLTNMGVYVSRHOL 321
 QY 370 KW-----NYEGFFYPSCTGSHLSRVCITYGAELFWLIKDGHWANFKDSYVDPLIK 423
 Db 322 -WVESECHNYMKD-----RS-CVFGIDVYDNLKMSALVAHKLYIESEBAFPC 368
 QY 424 L 424
 Db 369 L 369

Search completed: May 2, 2002, 07:48:48
 Job time: 213 sec

Fri May 3 10:57:56 2002

us-09-645-192-2.ispt

Page 9